DOCUMENT RESUME

ED 136 230 CS 003 288

AUTHOR Walmsley, Sean A.
TITLE The Criterion-Referenced Measurement of an Early Reading Behavior.
PUB DATE [76]
NOTE 103p.; Report prepared at the State University of New York at Albany; Appendix may be marginally legible

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; *Early Reading; Elementary Education; *Evaluation Methods; Measurement Techniques; *Norm Referenced Tests; *Reading Research; Reading Skills; *Reading Tests

ABSTRACT

In recent years, increasing criticism of norm-referenced measures of reading ability has sparked researchers' and educators' interest in criterion-referenced instruments. This study attempted to shed light on these issues by constructing a criterion-referenced test of early reading behavior, by presenting a rationale for its construction from both reading and statistical perspectives, and by administering this test to a stratified, random sample of school children in grades two, three, and four. Test results indicate that an understanding of the pronunciation of the consonant-vowel-consonant universe is well-established by the middle of grade two, that some further improvement occurs by the middle of grade four, and that a steady decline in recognition time occurs as a function of increasing age. Surprisingly, an almost equal number of errors were made on initial and final consonants. Results also suggest that the division of consonant-vowel-consonant groupings into various pronunication patterns (checked vowel, r-modified vowel, etc.) and into categories defined by whether the group is a real word or a nonsense word are both important dimensions of the consonant-vowel-consonant universe. (Author/KS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

#18

The criterion-referenced measurement of an early reading behavior

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Sean A. Walmsley
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

Dr. Sean A. Walmsley

ED136230

In recent years, there has been increasing criticism of norm-referenced measures of reading ability by educators and researchers, and a growing demand for criterion-referenced measures to replace them.[1] Attractive though the concept of comparing a child's reading ability with a specified criterion may be (in contrast to comparing it with his peers), there are many problems yet to be resolved. The most important of these is the apparent irreconcilability of a measurement technique on the one hand that demands the test's objectives to be precisely defined, and a subject matter (or process, depending on one's view of reading) that has continually resisted efforts to define it precisely. Also, even if such 'universes' may be defined, what is the relationship between a child's knowledge of that universe and the larger task of reading?

The construction of criterion-referenced measures and their item analysis also raises several issues. How should items be selected? Are traditional item generating procedures acceptable for criterion-referenced test construction? There are no generally agreed upon formulae for the determination of these and other important statistics.

The objective of this study was to shed some light on these issues by constructing a criterion-referenced test of an early reading behavior, presenting a rationale for its construction form both the reading and statistical perspectives, and administering this test to a representative sample of school children in the primary grades. A final stage was to analyze and interpret the test results, again from the two persepctives of reading and statistics.

The universe chosen for this study was the pronunciation of CVCs.[2] A rationale was presented for its utility in early reading, and it was fractionated into its constituent parts, or categories, from which a stratified random sample of items was generated to make up an experimental criterion-referenced measure.

The test was administered to a stratified random sample of school children in grades 2, 3 and 4; the results were analyzed and interpreted.

Test results indicate that an understanding of the pronunciation of the CVC universe is well established by the middle of Grade 2, for the students in my sample, and that while some further improvement occurs by the middle of Grade 4, mostly what happens between the three grades is a steady decline in the time it takes to recognize and pronounce CVCs. A surprising result however was obtained by analyzing the number of initial consonant, medial vowel and final consonant errors. Contrary to expectations, almost equal number of errors were made on initial and final consonants.

[1] Criterion-referenced measurement is the estimation of how much a student knows in terms of the degree to which his performance resembles a performance criterion; norm-referenced measurement is the estimation of how much he knows in terms of a comparison between his performance and that of others.

[2] The CVC is defined as consonant-letter, vowel-letter, consonant-letter; a written linguistic environment in which a dominant pronunciation pattern (the checked vowel--/æ/ as in bat) occurs.

CS 003 288

The results also suggest that the division of CVCs into various pronunciation patterns (i.e. checked vowel, r-modified vowel, etc.), and into categories which reflect the 'meaningfulness' (i.e. whether the CVC is a real word or a nonsense syllable; has high or low association value; etc.), are both important dimensions of the CVC universe.

The study's findings were also related to current issues in the construction of criterion-referenced reading tests.

Finally, the study made some proposals for the development of a final form of the CVC measure, suggesting the use of ordering theory, for example, for estimating hierarchies among items so that a shorter but no less representative test might be constructed for the CVC universe.

TABLE OF CONTENTS

Introduction..... 1

The Problem..... 6

Objectives of the Study..... 6

Review of the Literature..... 7

Issues in NRM and CRM
(a) Statistical perspective (item generation, item analysis, reliability, validity, reporting of scores)..... 7

(b) Reading perspective..... 13

Method: Test development, administration and analysis..... 21

(a) Test development..... 21

(b) Test administration and analysis..... 25

Results..... 28

Discussion..... 31

(a) Definition of the CVC universe and its categories..... 31

(b) Test construction procedures and their difficulties..... 34

(c) Interpretation of the results, from reading and statistical perspectives..... 36

(d) Procedures appropriate for development of final form of CVC test..... 40

INTRODUCTION

Improved techniques in educational measurement and changes in educational philosophy over the last fifteen years have raised some important questions about traditional forms of measuring and reporting student achievement (Glaser, 1963; Thorndike, 1971).

What are the primary means of determining student achievement? One is by estimating how much a student has learned, or knows, in terms of a comparison between his performance and the performance of others (Glaser, 1963; p. 520); this is usually referred to as norm-referenced measurement. The other is by estimating how much a student has learned, or knows, in terms of the degree to which his performance resembles a specified performance standard or criterion[1] (Ibid, p. 519); this is usually referred to as criterion-referenced measurement.[2]

The overwhelming majority of tests employed for educational purposes during the past fifty years have been norm-referenced in

---

[1] Although this definition is usually accepted in the literature for criterion-referenced measurement, it may be rather misleading. What is to prevent a test user specifying a criterion on a norm-referenced test? Typically, the criterion-referenced test is intended to determine how much of a subject matter a student knows, and the criterion represents a percentage of items correct rather than a student's standing in a population.

[2] Criterion-referenced measurement will be abbreviated to CRM, norm-referenced to NRM throughout the paper.

their design and use. In the area of reading instruction, norm-referenced measures are widely used for measuring students' reading status and growth (e.g. the California Reading Test, 1936-70; Metropolitan Achievement Test, 1931-71), diagnosing reading difficulties (e.g. the Stanford Diagnostic Reading Test, 1966-8; the Gray Oral Reading Test, 1963), selecting students for remedial reading programs, evaluating the effectiveness of programs, and so on.

The few available criterion-referenced measures of reading ability and achievement have, until recently, been largely informal (e.g. placement guides, end-of-book tests, etc). However, in the last two years, the major publishing companies have begun to publish comprehensive criterion measures in reading, principally in the areas of decoding and comprehension in the elementary grades, K-6 (Hawes, 1973; Thompson & Dziubau, 1973). There are now available criterion-referenced tests covering the broad spectrum of early reading skills (McGraw-Hill Prescriptive Reading Inventory, 1972; Cooper-McGuire Diagnostic Word-analysis Test for Word Attack Skills, McGuire-Bumpus Diagnostic Comprehension Skills, 1974), and for specific areas such as phonics generalizations (e.g. Domain Phonics Test, 1972; Sipay Word Analysis Test, 1974). Most of these new measures are tied to instructional "packages", which may be loosely characterized as "diagnostic and prescriptive" reading instruction; a few (e.g. the Domain Phonics Test) have been constructed for general testing purposes.

Considerable pressure has been brought to bear on the reading field to develop criterion-referenced measures of reading (Chall, 1967; Corder, 1971) and the rapid increase in CR measures during the last two years attests to a quick response to these demands. Why has the reading field become so concerned with criterion-referenced

measurement? Several trends suggest themselves as plausible hypotheses:

Firstly, it is possible that shifts in educational philosophy--in the direction towards education having as its primary emphasis the development of the individual (Bloom, 1971) have brought corresponding changes in educational measurement; consequently, measures that assess what an individual knows or does not know are preferred over those which estimate his general standing in a population.

A second, possibly more plausible hypothesis is that great strides have been made in the technology of measurement techniques (Cronbach, et. al., 1972), which, when coupled with more exacting demands made upon them by educators, evaluators and clinicians, have promoted a shift towards CRM--a more sophisticated and precise form of measurement in their view (Bissell, 1973; Bormouth, 1970; Corder, 1971).

Another hypothesis may be derived from recent political, social and educational demands for "accountability", "individualized instruction", "diagnostic and prescriptive teaching" and so on. Consequences, no doubt, of perceived declines in academic standards (e.g. Start & Wells, 1972; Northcutt, 1975) and increased costs in public education, as well as outgrowths of innovative educational programs, these have challenged test constructors to develop criterion-referenced measures which are anticipated to be found more effective in the evaluation of these innovations.

Finally, research in and articulation of "stages" in cognitive development (Piaget & Inhelder, 1969), in language development (Brown, 1973) and in other areas (e.g. Kohlberg, 1968) in recent

years have encouraged the construction of tests--mostly for research, but eventually for classroom use--that measure primarily not how a student compares with another student, but rather at what "stage" of development he is functioning. If stages in reading development are to be articulated in a manner similar to those in cognitive and language development (Chall, 1973), clearly there will be a need for measures appropriately constructed to determine a student's reading "stage". While these tests will always allow comparisons to be made between students' performances, their primary purpose will be to compare an individual student's performance with a specified standard; consequently, they are likely to be criterion-referenced.

A number of significant questions are raised by the introduction of criterion-referenced measurement into the field of reading instruction. Firstly, on what grounds may we be sure that criterion-referenced measurement can achieve what norm-referenced measurement has reportedly failed to achieve? In addition to the problem of major theoretical differences (i.e. test theory and construction), there are some important practical questions to be considered; for example, if norm-referenced measures are to be criticized on the grounds that they take their definition of reading from the performances of children on tests (see Bormuth, 1970), cannot criterion-referenced tests of reading be similarly criticized because they take their definition from "objectives", devised by teachers or test constructors? Do we have evidence that the latter approach adequately compensates for the weaknesses in the former? If norm-referenced test construction procedures are "subjective", are criterion-referenced test construction procedures necessarily "objective"?

Secondly, are the statistical procedures for the development of CR tests in reading (i.e. item selection, item analysis, reliability and validity procedures) sufficiently well established that we may be confident of the reliability and validity of these new measures?

Thirdly, has the reading process (particularly as it relates to reading comprehension) been sufficiently well defined that the basic underlying assumptions of CRM (i.e. that the test's "domain" or "universe"[1] be precisely defined; that items be drawn randomly from pools of all possible items; that item pools represent unambiguous categories; and so on)may be met with confidence?

The developers of criterion-referenced tests appear either not to have considered these questions at all, or do not have many reservations about the answers to them. One reason for this may lie on the fact that two somewhat separate groups are involved in the development of reading tests--researchers from the field of statistics and educational measurement, and researchers from the field of reading theory and instruction. While these two groups may be aware of the developments in their own field, are they always conversant with developments in each other's fields? Are criterion-referenced test in reading the result of a partnership in which there is equal accessibility to the accumulated knowledge of both fields?

---

[1] A "universe" may be defined as a set of conditions under which a subject might be observed (Cronbach, et. al., 1972; p. 9); it implies a comprehensive set of conditions that all belong to the same set.

THE PROBLEM

The questions posed above have not received substantive consideration in the research literature. Nor does there appear to be evidence relevant to these questions from the empirical use of the measures. In fact, the research on criterion-referenced measurement appears to be so limited that one wonders why the major publishing companies acted so quickly to produce them. Clearly the educational considerations were strong, as noted earlier. It is unfortunate, though, that some essential preliminary research was not conducted before the wide-scale production of CR tests in reading was undertaken.

Thus, while increasing demands have been made on the reading field to develop criterion-referenced measures of reading ability and achievement, and in response to these demands a number of CR reading tests have been published, some important questions remain unanswered. There are questions from a statistical perspective that ask, for example, what procedures should typically be employed to generate items from specified categories, to analyze items to ensure their representativeness, and to determine the traditional co-efficients of reliability and validity. There are questions, too, from a reading perspective. For example, how does one define a "domain" in reading? What are the reading "domains"?

OBJECTIVES OF THE STUDY

The study was intended to shed some light on these two perspectives (i.e. reading and statistical) of the criterion-referenced measurement of reading. In order to address the question of defining a "domain" in reading, the author chose a domain (pronouncing trigrams

of the CVC pattern[1]) and presented a rationale for its choice. In order to deal with some of the problems inherent in devising a criterion-referenced test, the author constructed a test of this CVC "domain", following accepted procedures for CR test construction. In both these tasks, the purpose was to show what kinds of difficulties arise in criterion-referenced test construction in reading, and to assess various means of overcoming the obstacles.

Since much of the criticism of criterion-referenced measures has been directed at their administration and interpretation, the author additionally sought to address these issues. Consequently, the completed CVC pattern test was administered to a representative sample of school children, and the results analyzed.

Again, the purpose of this administration and interpretation was to shed light on the various difficulties associated with each stage of test construction, and so the interpretation of the results was made from both the reading and statistical perspectives.

REVIEW OF THE LITERATURE[2]

Issues in norm-referenced and criterion-referenced measurement:

(a) Statistical perspective

While CR measures are proposed to compensate for the alleged inadequacies of NR instruments (see Corder, 1971; pp. 58-9), there are some formidable knots yet to unravel. Some authors (e.g. Davis, 1972; Ebel, 1972; 1973) debate whether CMR is really different from NRM save

---

[1] The CVC pattern is defined as consonant-letter, vowel-letter, consonant letter; a written linguistic environment in which a dominant pronunciation occurs (i.e. the checked vowel--/æ/ as in bat).

[2] This review is necessarily brief: a more detailed version may be found in Walmsley (1975; p 13 ff).

in terms of the reporting of scores. An examination of the literature and of the test construction procedures for NRM and CRM point to some well-established differences, but it also reveals how much more research is needed to develop appropriate test construction an item analysis procedures for CRM (Walmsley, 1973).

Questions to be resolved include those surrounding item generation procedures (Hively, et. al., 1968; Kriewall, 1969; Bormuth, 1970); item analysis procedures (Cox & Vargas, 1966; Brennan, 1970); the determination of reliability (Livingston, 1972) and validity coefficients (Osburn, 1968; Dahl, 1971); and the reporting and interpretation of test scores (Millman, 1973).

1. Item generation procedures.

Items for NR measures typically are selected on the basis of their representativeness of the test's objectives, and are generated either by arbitrary choice of the test constructor (Gronlund, 1968) or by use of a content-by-behavior table of specifications (Bloom, et. al., 1971). Usually many more items are generated than are used, the surplus being eliminated by item analysis procedures. This technique has been criticized for its arbitrary and non-rigorous selection of items (Bormuth, 1970).

For the reason that the purpose of a CR measure is to estimate how much a student knows the subject matter tested, items have to be generated according to a much stricter set of procedures. Resting on the base of generalizability theory[1] (Cronbach, et. al., 1972),

---

[1] Generalizability theory rests upon the mathematical model of probability theory. Thus, if one constructs a test in which items comprise a random sample from a defined universe of items, then a subject's performance on the sample of items may be generalized to an estimate of his performance on the whole universe (see Cronbach, et. al., 1963; 1972).

a criterion measure generates its items--theoretically, at least--by randomly sampling from all possible items within each "universe" or "domain".

Thus the item generating procedures for NR measures differ quite substantially from CRM, although this is to be expected given their different emphases in purpose. But the use of a random sampling scheme as an alternative to arbitrary item generation procedures raises some difficulties of its own. One of these is specifying unambiguous categories from which the total item pool can be constructed; another is generating all possible items. It may be found, though, that the problems inherent in these procedures have less to do with measurement theory than with the conceptual difficulty in fractionating complex psychological phenomena. It would not be difficult to generate items randomly from item pools were it not difficult to define item pools in the first place.

2. Item analysis procedures.

The two principal parameters in traditional item analysis are item difficulty and item discriminating power. Item difficulty usually refers to the proportion (or a transformation of this proportion--i.e. a "z-score") of subjects scoring an item correct. Item discriminating power refers to the correlation between an item score and the total test score (Henrysson, 1971; Ebel, 1972). Item difficulty indices are computed to permit the ordering of items in ascending levels of difficulty, and to allow the test constructor to specify the range of difficulty amongst items. Item discrimination indices are computer primarily to select those items which maximally discriminate between subjects' performances. Both these procedures are

essentially norm-referenced concepts--an item's difficulty (as measured by traditional procedures) is a function not only of the item's inherent complexity but also of the ability of the subject. The selection of items of medium difficulty and high discriminating power may ensure maximum variance in student scores, but the items thus selected need not necessarily be those that most adequately represent the universe (Hopkins & Wilkerson, 1965).

Thus it may be found that item discriminating indices are of little utility for a CR measure. Alternatives have been suggested--for example, Dahl (1971) has argued that the proper concern of CR item analysis should be "objective-item" congruence, and "item homogeneity" (Hively, et. al., 1968); in other words, determining the relationship between items and the test's objectives, and between the items themselves.

If the purpose of the test is primarily to estimate the relative standing of a student amongst his peers with respect to some subject matter (i.e. NRM), then item difficulty and discriminating indices are especially useful. However, if the test's purpose is to determine how much or what a student knows of a subject matter (i.e. CRM), irrespective of his peers, then the items must adequately represent the subject matter, and the purpose of item analysis is to show that they do indeed achieve this. It is clear that traditional procedures are not well suited to this latter purpose, but it has yet to be resolved whether modifications to existing procedures (Cox & Vargas, 1966) or radically new procedures (Rasch, 1966; Bart & Krus, 1973) will be found effective.

3. Reliability.

There is some debate in the literature over the reliability[1] coefficients for CR measures (Livington, 1970, 1972; Ivens, 1970; Ozenne, 1971; Harris, 1972; Shavelson, et. a., 1972). This controversy raises some important issues, and suggests that different forms of reliability present different problems for CRM. Neither inter-rater nor internal consistency are troublesome, the determination of the former lying outside the test itself, and the latter being dependent on the "unifactorial" nature of the test--something a CR measure strives for. However, both test-retest and parallel forms reliability depend on the variance in subjects' scores: a variance actively sought in NRM, but incidental to CRM. Thus it may be that traditional procedures are not so useful for CRM. An alternative might be to employ factor analysis to determine "item homogeneity"--if all the items measure the same thing, then they should continue to measure it whether given twice or ten times (test-retest), or split into two or three alternative forms (parallel forms), or divided into two halves of the same test (internal consistency). Procedures such as factor analysis do not depend on variance in subjects' scores, and offer an alternative for estimating the repeatability of criterion-referenced measures.

---

[1] We may identify four types of reliability--test-retest, parallel forms, inter-rater, and internal consistency. Test-retest is calculated by correlating scores for the same group of subjects on two administrations of the samo test; parallel forms is calculated by correlating the scores on one administration of the two (or more) forms; inter-rater by calculating correlations of independent judgements of the same test; internal consistency by correlating one half of the test with toher (Stanley, 1971; Nunnally, 1967).

4. Validity.

If the validity[1] of an instrument is a measure of its ability to do what it was intended to do (Nunnally, 1967), then differences between NRM and CRM in determining validity coefficients will essentially reflect the different purposes of the two means of measuring student performances. NR measures primarily serve selective and predictive purposes, thus their validity will be determined primarily by the efficacy of their selection and prediction. CR measures are primarily intended to sample universes, thus their validity will primarily be determined by establishing the adequacy with which items represent universes. If NR measures tend to stress concurrent and predictive validity, while CR measures emphasize content validity, it is only an appropriate reflection of their different emphases in purpose.

The procedures for establishing the validity of criterion measures have not been generally agreed upon. Some ingenious suggestions have been made (e.g. Cox, 1971), but most researchers agree with Hambleton & Gorth (1970) that the validation of criterion measures will require both correlational and non-correlational procedures.

5. Reporting and interpreting scores.

In NRM, scores are reported in terms of comparisons between students, typically as percentiles, grade- or age-equivalents. In CRM, scores are reported in terms of how much or what a student knows,

---

[1] We may identify four principal kinds of validity: concurrent (i.e. an independent judgment at the same time should give the same results); predictive (i.e. a test score should correlate with a score on another measure sometime in the future); content (i.e. the test should measure what it says it measures); and construct (i.e. the test should bear out a theory underlying the trait being measured (Cronbach, 1971; Nunnally, 1967).

typically as criterion or mastery scores. A score of 85% in an NR test means that a student has learned (or knows) the subject matter covered by the test better than 85% of the other students taking the test; a score of 85% on a CR measure means that a student has learned (or knows) 85% of the subject matter itself--irrespective of the performances of others.

NR scores have been criticized because they do not say what the student knows or does not know; but what is the meaning of a CR score? At what level should a mastery score be set to establish that the student "knows" the subject matter? Bayles (1934) has commented that: "the child does not go from a condition of no understanding to a condition of complete understanding. He goes instead from a condition of lesser understanding to a condition of greater understanding" (p. 205). In CRM, there needs to be firstly sufficient items to represent the universe, and secondly the establishment of the number of items correct which confidently demonstrates the student's "greater" or "lesser" understanding. Thus the setting and interpretation of the mastery score will depend on the degree of certainty the test user wishes to establish for the adequacy of the measure in these two respects. Millman (1973) has set out some useful guidelines for this determination, but the replacement of percentiles and grade equivalents with criterion and mastery scores raises a number of important questions that have so far not received unequivocal answers.

(b) Reading perspective.

The development and administration of NR reading measures have not presented test constructors with unsurmountable problems. In fact since the construction of the first NR tests in reading (Thorndike,

1914) there has been a steady growth in the use and technical sophistication of these measures.

Since NR measures do not require strict definitions of reading universes, their construction has never been hampered by having to define what constitutes the content and process of reading, since the preformances of students on tests determines what finally appears as the test content. These tests continue to be effective in the task for which they were created, namely comparing the reading ability/achievement of students.

CR reading tests on the other hand raise some questions for which there are as yet only tentative answers: what is the content of reading? In what ways does the reading process fractionate? Can universes in reading be defined unambiguously? How does one articulate what reading consists of without reference to the reading performances of students?

Most developers of CR reading tests have avoided a direct confrontation with these questions by adopting definitions of reading based mainly on the content of existing test or reading curricula. One strategy is to examine the major reading series and to create groups of reading "objectives" from which items were derived, tried out on a representative sample of students, analyzed and revised to make final tests. In fact such tests (e.g. Prescriptive Reading Inventory: McGraw-Hill, 1972) differ only very slightly from NR instruments in their test constructional procedures, though the reporting of results is radically different. The definition of reading in these tests is really no different from the definitions adopted by constructors of NR tests: both types of test are derived from essentially the same body of research on the

components of the content and process of reading.

To return to the question, can reading be fractionated? While identifying the subskills in beginning reading appears an easier task than in later reading, researchers have not agreed on the definition or the ordering of the subskills in beginning reading with any more certainty than they have for later reading skills. Indeed, there is a growing suspicion that these definitions and hierarchies will continue to elude us (NIE Conference on Reading, 1974). The definition of comprehension skills has also been relatively unsuccessful, perhaps for the reason offered by Simons:

> "The reason for the lack of progress in comprehension goes much deeper than poor research methodology. The current lack of descriptions of the mental processes involved in reading comprehension render it very difficult to establish adequate behavioral criteria for successful comprehension. This is in turn due to comprehension being a covert mental process which...takes place without any overt behavior being produced." (1970: p. 354)

Other researchers (Miller, 1973; Smith, 1975; Thorndike, 1973) have criticized traditional definitions of comprehension (e.g. Davis, 1944) and suggested that reading comprehension is not easily broken into constituent parts.

What are the implications of these theories? Firstly, CR tests derived from definitions of reading that are essentially the same as those used in NR tests are unlikely to be substantially different from the NR tests they are intended to supplement or replace. Secondly, current research in reading appears to be suggesting that the fractionation of reading (both decoding and comprehension) into constituent, perhaps even hierarchical components is very much more difficult than was previously thought.

As Mann & Phillips (1967) suggest:

> "The basic assumption underlying the development and utilization of fractional approaches is that human behavior may be successfully separated, as it were, into specific entities, units or functions, these being essentially independent and capable of being indiviudally evaluated and/or exercised. Advocates of fractional approaches appear to be proceeding as if this assumption were a fully demonstrated and effective operational viability" (p. 312).

These reservations about the feasibility of defining reading universes have important implications for the future development of criterion-referenced measurement.

Rationale for the choice of the CVC pattern as a universe in reading

One of the first tasks in learning to read in one's own language is to learn "a printed code for the speech we possess" (Chall, 1967). While approaches to beginning reading instruction vary considerably, specific instruction in learning this code is an internal part of the first grade experience.

Research in beginning reading has established that although there are some consistent relationships between written symbols--graphemes--and their sounds--phonemes--(Weir & Venezky, 1966), mostly "spelling-to-sound rules are based not upon (individual) letters or graphemes as such but rather upon functional spelling units" (Venezky, 1966, p. 83). Fries (1963) refers to these units as spelling patterns which "represent the word patterns; but the individual letters in the sequence of a spelling pattern do not necessarily duplicate the sequence of the phonemes in the word pattern for which the spelling pattern stands" (p. 169).

Research into the utility of these spelling patterns (i.e.

whether the use of the generalization results in the correct pronunciation or recognition of a word) had tended to cast some doubt on their usefulness. Clymer (1963) examined 45 "phonic"[1] generalizations in everyday classroom use, and determined that only 18 of them were sueful. Subsequent studies by Bailey (1967), Emans (1967), Davis (1972) and Emans & Harms (1973) have generally confirmed Clymer's findings, namely that there are a somewhat limited number of phonic generalizations that are highly applicable to phonics instruction.[2]

Of these generalizations, all studies attest the usefulness of the CVC pattern.

The CVC pattern[3] may be viewed in two ways. Firstly, it may be regarded as a phonic generalization--"when a vowel is in the middle of a one-syllable word, the vowel is short" (Clymer, 1963). In this case, the pattern "CVC" refers to a consonant-sound, vowel-sound, consonant-sound, and may be represented by a consonant-letter, vowel-letter, consonant-letter (i.e. CVC-bat), or by other combinations (i.e. VC-at; CVCC-camp; CCVC-trip; CCVCC-tramp). The utility of this generalization, stated in this way, has been found empirically to be 62% (Clymer, 1963); 73% (Emans, 1967); 71% (Bailey, 1967); 66% (Davis, 1972). Variations in utility ratings result primarily from differences in word corpora.

---

1 The term phonics refers to a pedagogical device for teaching sound/symbol relations.

2 Studies use slightly differing criteria for determining "usefulness". For a discussion, see Walmsley (1975) p. 31 ff).

3 For ease of understanding, CVC meaning "consonant-sound, vowel-sound, consonant-sound" will be referred to as "CVC"; CVC meaning consonant-letter, vowel-letter, consonant-letter will be referred to as CVC.

Secondly, the CVC pattern may be seen as a written linguistic environment in which the vowel is pronounced 'short' or 'checked' (i.e. CVC, /bæt/). In this case, the CVC refers to a spelling pattern, and the utility of such a pattern would be expressed in terms of how many times one might find words containing or comprising this pattern, pronounced regularly with a checked vowel. The tullity of this generalization, when stated this way, has been found empirically to be 92% (Emans & Harms, 1973). This percentage of utility is as high for the environment CVC as it is for the combined utilities of other checked vowel environments (i.e. CCVC, CCVCC, etc). Thus "CVC" as represented by CVC may be said to be a very common environment in which the checked vowel phenomenon occurs.

It would be tempting at this point to derive the CVC universe from all possible combinations of CVC were it not for the fact that not all of these combinations represent the dominant pronunciation pattern. Some combinations (for example, CVr, CVy, CVw) represent different pronunciation patterns (e.g. bar, day, cow). Others represent irregular pronunciation patterns (e.g. guh, xyz). Most CVCs are pronounced regularly, some having both regular and alternative pronunciations, and a few are either difficult to pronounce or are unpronounceable according to the usual rules and principles of English phonology. Nevertheless, all CVCs can be generated, and most of them either exist as sequences in words in the language, or are capable of being easily pronounced. There is, of course, no reason why any CVC should not at some time have meaning or pronunciation attached to it; indeed, many CVCs which we refer to now as "nonsense syllables" were at one time in common usage (e.g.

flp--a fivepenny piece), and some CVCs are common words of only recent vintage (i.e. daz, mod).

The CVC universe chosen for this study was all possible combinations of CVC (See Table 1)[1]. The definition of the universe is, however, only one aspect of the task of constructing a test for the CVC pattern. A second aspect is concerned with how a student demonstrates knowledge of this universe; and a third, with the value or utility to be derived from such a knowledge in terms of the larger task of reading.

There are several ways in which a student may demonstrate knowledge of the CVC universe. He may do it as he read; or as he responds to the task of explicitly pronouncing CVCs in a test; or as he describes the principle of a checked vowel in a CVC as a phonics generalization. In the first of these tasks--reading--the demonstration itself may not be directly observable as in the case of an experienced reader reading silently. In the last of these--saying the "rule"--we cannot assume that an accurate description of the generalization necessarily is to be equated with the knowledge of the CVC universe itself. For the purposes of the present study, the second approach--explicitly pronouncing CVCs--was adopted as the means by which knowledge of the CVC universe was to be estimated.[2]

---

[1] Even the definition CVC raises difficulties: which letters are consonants? Some letters sometimes represent consonants and sometimes vowels (e.g. h, w, y have consonant sounds in the initial position but rarely in the final: rather they modify the medial vowel and are pronounced with that vowel sound as a unit). The letter r usually has a consonant sound in the final position, but it, too, modifies the medial vowel sound. The letter y additionally represents a vowel sound in its own right.

[2] An important factor here is the speed with which a student is able to pronounce the CVC. Because of well-established evidence of decreases (with age primarily) in reaction and recognition times in cognitive tasks (White, 1970), latencies between stimulus and response were recorded in this study.

TABLE I

Definition of CVC (consonant-letter, vowel letter, consonant-letter)

INITIAL CONSONANT	MEDIAL VOWEL	FINAL CONSONANT
b, c[1], d, f, g[2], h, j, k, l, m, n, p, q[3], r, s, t, v, w[4], x[5], y, z.	a, e, i, o, u, y.	b, c[1], d, f, g[2], h[6], j, k, l, m, n, p, q, r[7], s[8], t, v, w[4], x, y[9], z.

[1] c has two pronunciations, /k/, /s/.

[2] g has two pronunciations, /g/, /ǰ/.

[3] qVC never occurs in the English language.

[4] w modifies the medial vowel (a, o) in the initial position; it has no sound as a final consonant-letter; when w is in the final position, the vowel is diphthongized. *

[5] x has the sound /ks/ but is usually pronounced /z/ when it occurs in the initial position of a syllable containing more than the letter x (i.e. xerox).

[6] h has no individual sound as a final consonant-letter; the combinations ah, eh, oh, uh are lengthened.

[7] r is sometimes (according to dialect) not pronounced as a consonant in the final position; here, the r may affect the vowel quality.

[8] s has two pronunciations, /s/, /z/.

[9] y does not usually (except in certain dialects) represent a consonant sound in the final position; when y is in the final position, the vowel is diphthongized.

* I include under diphthongs the following: /iy/; /ey/; /ay/; /aw/; /ow/; /uw/; /oy/.

The nature and extent of a student's knowledge of the CVC universe should not be confused with the utility of his understanding. That a student can pronounce CVCs, or say the rule, does not tell us much about the usefulness of this knowledge as he learns and uses the printed code. Since this particular question has (to the author's knowledge) yet to be resolved empirically, we may only speculate on the utility of CVC knowledge.

If 92% of all "C-checked V-C" patterns are represented by the written environment CVC, then we may suppose that a knowledge of at least part of the CVC universe (i.e. those CVCs which are pronounced regularly) will assist the student correctly pronounce monosyllabic words with a medial checked vowel. Furthermore, such a knowledge may be useful in distinguishing between words that follow the generalization (e.g. bat) and those that do not (e.g. day). Finally an understanding of the CVC universe may be of help in pronouncing and recognizing longer words in which CVC patterns are embedded (i.e. ped in impediment).

Eventually, though, the skilled reader needs to have internalized the CVC pattern and "automatized" it to the point at which the understanding of the pattern is implicit knowledge, instantly available whenever needed.

It can be seen from this discussion that even if we choose a reading domain whose general parameters can be fairly well described, many judgements still need to be made about what constitutes the domain and what lies outside it. The CVC universe could have been defined several ways. Also, there are many kinds of response behaviors available for the task of defining what is meant by knowledge of the CVC universe--e.g. pronouncing CVCs, stating the phonic

generalization, etc. The definition of the CVC universe and the response behavior to be associated with demonstrating knowledge of that universe are essentially arbitrary decisions made by a test constructor, and represent just one way of approaching the criterion-referenced measurement of this early reading behavior.

METHOD: TEST DEVELOPMENT ADMINISTRATION AND ANALYSIS

(a) Test Development

The first step is to define unambiguous categories subsumed by the universe, since it is clear from the preceding discussion that the CVC universe does not comprise a single category.

The definition of categories is necessary for two reasons: firstly, there are subdivisons of the CVC universe that arise from the structural and phonological differences between various CVCs. These include the pronunciation patterns (i.e. regular checked vowel, r-modified vowel, etc). Secondly, the child's ability to pronounce a given CVC is much influenced by the experience he may have had in pronouncing it (in the case of real word CVCs), or how closely the CVC (in the case of a nonsense syllable) resembles a real word. The child's experience with a given CVC may be dependent on several factors: (1) whether the CVC is a real word or a nonsense syllable; (2) its frequency as a real word; (3) its frequency as a syllable. Consequently, a number of categories were established to control for these factors.

Pronunciation patterns

Initially, we may divide the CVC universe into its various pronunciation patterns (see Table 2). There are six such patterns: the checked vowel, the r-modified vowel, the w-modified vowel, the y-modified, the h-modified, and irregular pronunciation patterns.

Lexical value

Next, we may divide the CVC universe into categories that represent the lexical values of CVCs. There are two such categories proposed here: (a) real words; (b) nonsense syllables.[1]

Frequency value

A third division is into categories which represent the range of frequency values of CVCs. The frequency value is a reflection of the number of opportunities for seeing or using the CVC in reading materials available to the beginning reader. The frequency index may be estimated in several ways. For this study, two were chosen. Firstly, it was determined whether a CVC existed as a word at all; then its frequency in materials up to the fourth grade was calculated. Secondly, an estimate was made of the CVC's frequency as a syllable in polysyllabic words, both in materials appropriate to elementary school, and in

[1] Definitions of these will be found in Table 3.

TABLE 2

Dimensions of the CVC universe: 1-definitions of categories by pronunciation patterns[1]

1. Regular (with alternatives)	CVCs containing a medial checked vowel, pronounced either regularly [/æ/ as in bat], or modified [/ɑ/ as in was; /ʌ/ as in won]
2. R-modified	CVCs in which final r modifies the medial vowel, and is usually pronounced in combination with it [/ɹ/ as in bar; also /ɛ/ as in ber'ry]
3. W-modified	CVCs in which final w modifies the medial vowel, and is pronounced in combination with it [/o/ as in low; /ɔ/ as in saw]
4. Y-modified	CVCs in which final y modifies the medial vowel, and is pronounced in combination with it [/ey/ as in day; /iy/ as in key]
5. H-modified	CVCs in which final h modifies the medial vowel, and is pronounced in combination with it [/ɑ/ as in bah; /e/ as in heh]
6. Irregular	CVCs in which the following elements occur: aq; eq; iq; oq (except coq); uq; qa; qe; qy; qi; qo; qu; ih; oh; uh (except yuh); yh; iw; uw; yw; iy; uy (except buy, guy); yy

[1] Phonetic transcriptions here are approximate only; pronunciations vary from dialect to dialect.

TABLE 3

Dimensions of the CVC universe: 2-definitions of categories by lexical, frequency values, etc.

-category-	-definition-
(A) REAL WORDS	[having an entry in Webster's International Dictionary, 1972]
Frequency as a word	
-rare, obsolete, slang	[defined as such in Webster's Dictionary]
-low	[0-1 actual frequency in reading materials up to 4th grade, taken from Carroll, et. al., 1971]
-middle	[2-26 actual frequency]
-high	[26-6900 actual frequency]
Frequency as a polysyllable	
-does not exist	[does not exist as CVC in any word in Webster's Dictionary][1]
-exists	[exists as CVC in a word in Webster's Dictionary, but does not occur in 'low' category]
-low	[occurs once as polysyllabic syllable in word corpus,[2] from Berdiansky, et. al., 1969]
-middle	[occurs 2-4 times in Berdiansky, et. al.]
-high	[occurs 5 or more times in Berdiansky, et. al.]
Pronounceability Value (see Table 3-8)	
-low	[sum of individual letter pronounceability values equals more than 6]
-middle	[sum of individual letter pronounceability values equals between 4 and 6]
-high	[sum of individual letter pronounceability values equals less than 4]

[1] According to the author's review (see Appendix [A]).

[2] This list of polysyllabic syllables was derived from Berdiansky's corpus of 8,600 words.

TABLE 3 (contd)
Dimensions of the CVC universe: 2-definitions of categories by lexical, frequency values, etc.

-category-	-definition-
(B) NONSENSE SYLLABLES	[not having an individual entry in Webster's International Dictionary, 1972]
Association Value:	[as for real words]
-does not exist -low -middle -high	
Pronounceability Value:	[as for real words]
-low -middle -high	
Frequency as a polysyllable:	[as for real words]
-does not exist -exists -low -middle -high	

words in general.[1]

Association value

Another category subsumed within the CVC universe is one which reflects the closeness of a CVC to natural language. In one sense, the division of CVCs into real words and nonsense syllables is one dimension of meangfulness; the association value however reflects how closely to real words nonsense syllables are perceived to be. No association vlaues have been calculated for children, but there are several adult association value lists available (Glaze, 1928; Krueger, 1934; Underwood & Schulz, 1960; Archer, 1960) and for the purpose of this study, Archer's association values were used on an experimental basis.

Pronounceability value

Finally, a category was created which was intended to reflect the ease of difficulty of pronunciation of CVCs. It is not possible to create a category for pronunciation which excludes the effects of frequency and association, and the task is hampered further by the fact that no list of CVCs by this dimension has been produced (to the author's knowledge). An experimental set of values for pronounceability was therefore constructed on the basis of a combination of the frequency of individual letters in various positions (Thorndike & Lorge, 1944) and the consistency of phonemic correspondence (Venezky, 1963).[2]

---

[1] This last estimate was achieved by examining words in Webster's Dictionary to determine whether CVCs existed in words other than those found in elementary reading materials.

[2] For a detailed discussion of this procedure, see Walmsley (1975; pp 52-5).

Test construction procedures

Having generated all possible combinations of CVC (see Table 1), these 2646 combinations were categorized according to the dimensions referred to above: real words/nonsense syllables; frequency as a word; frequency as a polysyllable; association value, pronounceability value. In addition, they were categorized according to their pronunciation pattern. In this way, a table of specifications was constructed; it may be seen in Table 4(a) and 4(b). It will be noted that some items appear in more than one category, which is a function of categories not always being mutually exclusive.

From this table of specifications a stratified random sampling plan was prepared, in order to generate items in samples proportionate to the size of item pools. A sampling ratio of 1:10 was adopted. This ratio yielded a total sample of 914 items for the complete CVC test. In order to be able to administer the test in manageable units without losing any generalizability, a procedure known as multiple matrix sampling (Shoemaker, 1972) was employed; The technique involves giving random samples of items to random samples of subjects, and combining the results in such a way as to allow estimates to be made about the population of students' performances on the population of items. The sampling of items and students met the requirements for multiple matrix sampling (i.e. stratified random sampling), and the procedure was adopted for the study.

Test items were randomly generated from the item pools. Next, items were randomly assigned to one of six subtests, each comprising 152 items. Finally, the order of items (1-152) was randomized, so that within each subtest of 152 items, the original categories were

TABLE 4 (a)
Dimensions of the CVC universe: 3- table of specifications

REAL WORDS	regular	r-mod	w-mod	y-mod	h-mod	irregular
Frequency as a word:						
-rare, obsolete, slang	223 (rep)	23 (vir)	12 (zow)	10 (hoy)	11 (sah)	0 (-)
-low	110 (jab)	4 (cur)	9 (yew)	5 (guy)	3 (hah)	0 (-)
-middle	98 (mix)	2 (bar)	9 (paw)	3 (joy)	2 (yah)	0 (-)
-high	96 (had)	10 (fur)	10 (how)	15 (way)	0 (-)	0 (-)
Frequency as polysyllable:						
-doesn't exist	45 (fez)	0 (-)	2 (yew)	3 (guy)	10 (huh)	0 (-)
-exists	171 (vat)	7 (sir)	17 (caw)	9 (bay)	3 (pah)	0 (-)
-low	111 (din)	2 (ger)	9 (how)	8 (coy)	2 (dah)	0 (-)
-middle	121 (don)	8 (her)	7 (jew)	10 (pay)	1 (jah)	0 (-)
-high	79 (beg)	22 (per)	5 (low)	3 (day)	0 (-)	0 (-)
Pronounceability value:						
-low	1 (wag)	0 (-)	5 (gaw)	5 (jay)	0 (-)	0 (-)
-middle	86 (gus)	0 (-)	35 (cow)	28 (key)	6 (nah)	0 (-)
-high	440 (bad)	39 (jar)	0 (-)	0 (-)	10 (lah)	0 (-)
Association value:						
-does not exist	28 (tut)	0 (-)	1 (wow)	0 (-)	3 (-)	0 (-)
-low	10 (yob)	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)
-middle	51 (het)	3 (ler)	3 (taw)	1 (gey)	8 (reh)	0 (-)
-high	438 (tin)	36 (tar)	36 (saw)	32 (may)	5 (yah)	0 (-)
TOTALS	527	39	40	33	16	0

Key: Number= # of items in item pool; (CVC)= example of CVC in category

TABLE 4 (b)
Dimensions of the CVC universe: 3- table of specifications

NONSENSE SYLLABLES	regular	r-mod	w-mod	y-mod	h-mod	irregular
Frequency as polysyllable:						
-does not exist	839 (buj)	22 (xur)	14 (naw)	15 (xoy)	9 (vah)	448 (fyy)
-exists	347 (biv)	28 (hir)	5 (gow)	9 (cey)	2 (feh)	7 (lih)
-low	103 (nec)	8 (jor)	1 (kew)	3 (noy)	0 (-)	1 (req)
-middle	71 (liv)	12 (zor)	0 (-)	2 (sey)	0 (-)	0 (-)
-high	34 (hol)	11 (ter)	0 (-)	0 (-)	0 (-)	0 (-)
Pronounceability value:						
-low	75 (xug)	0 (-)	10 (xew)	11 (zay)	2 (xah)	136 (qyq)
-middle	552 (jax)	6 (xur)	10 (faw)	18 (loy)	5 (gah)	275 (biw)
-high	767 (hos)	75 (cir)	0 (-)	0 (-)	4 (feh)	45 (hih)
Association value:						
-does not exist	83 (beb)	7 (rar)	2 (waw)	3 (yay)	0 (-)	38 (wyy)
-low	460 (wyj)	14 (zyr)	6 (zew)	8 (zey)	4 (xah)	283 (kyw)
-middle	476 (tul)	25 (zer)	8 (jow)	12 (vay)	6 (leh)	133 (hyq)
-high	375 (wac)	35 (bir)	4 (naw)	6 (sey)	1 (mah)	2 (liq)
TOTALS	1394	8i	20	29	11	456

Key: Number= # of items in item pool; (CVC)= example of CVC in category

represented, but not in any order. This same randomized order of items was repeated for each of the six subtests. For example, category R101 (Real Words by Frequency Value: rare, obsolete, slang) comprised 24 items (drawn randomly from the original item pool of 223 items). These 24 items were randomly assigned to six subtests, each comprising 4 items from this category. Next, these 4 items were randomly assigned to positions between 1 and 152 (in this case, to positions 90, 113, 119, 146).

(b) Test administration

The six subtests of 152 items each were prepared as slides for projection (35 mm slides: 10pt typeface in lower case pressed onto an acetate base with transfer lettering). When projected, each CVC measured 1 1/2" x 3/4" on a table viewer, inside a mask whose outside dimensions were 4" x 2".

To record latencies between stimulus and response, the slide projector lamp was wired through an on-off switch which actuated a timer (pulse generator, 1/10 sec). This timer was controlled by a footpedal and actuated when the subject's response was given. Since reaction times were subject to vary between tests and occasions, the latencies recorded by the procedure were necessarily relative not absolute estimates.

The test items were divided into two "trays" (75 slides in each). Two testers (the author and a colleague) were used. The actual task consisted of two sessions in which the subject was presented with 76 slides of CVCs to read. The slide would appear, the child would read it, the lamp would be turned off, verbatim response recorded, slide changed, and so on. A standard procedure was used to introduce the

subject to the task, and some practice items were given. The two sessions were separated by approximately one week. Order and practice effects were controlled by giving half the items forward, the other half backward.

A record of the actual responses was made by the tester. A sample of subjects were tested by both examiners to estimate inter-rater reliability.

All testing was carried out individually in schools where subjects were enrolled. In all cases, testing was done in quiet, distraction-free areas, with adequate heat and light.

Selection of subjects

Subjects were drawn from two elementary school systems within 40 miles of Boston. From complete registration lists of all students in three grades (2,3,4) 24 subjects per grade per school system were selected randomly. Subjects were assigned randomly to one of six subtests, and alternate subjects were given the test forwards or backwards. Testing was done in two sessions over a period of one month, alternating between the two school systems to control for time or practice effects.

Additional data gathered from students

The following additional data was gathered from students: (a) age (b) grade (c) sex (d) reading group within grade (e) teacher's estimation of the student's reading ability, his level of phonics knowledge, and what he might be expected to do on a CVC test.[1]

[1] Teachers' estimations were gathered by questionnaire.

Scoring the item responses

Item responses were coded and transformed into item scores (for the whole CVC-all elements had to be pronounced correctly), and scores for individual elements (initial consonant, medial vowel, final consonant).

Each item response also received a latency score, consisting of a figure representing in tenths of seconds the time which elapsed between the presentation of the slide and the student's response to it.

Groupings of the students

Two major groupings of the students were made for the purpose of the test analysis and interpretation. The first was by grade; the second by reading ability within grade (obtained from the teacher questionnaire).

Combining the scores through multiple matrix sampling

As discussed earlier, Shoemaker's multiple matrix sampling procedure was used to allow a total of 914 items to be given to 144 students without every student responding to all 914 items. Samples of items were given to samples of students, so that all students responded to some items within each category. In this way, an estimate could be made of what all students would have scored had they taken all items. In the present study, some modifications to the procedures were made in the analysis of the results--for example, the procedure is not intended for cases where there are less than two items per subtest. Since items were randomly generated from strictly defined categories, and since subjects were randomly selected from within

grade, it was felt that items within categories and subjects within grade were equivalent, and the means for items within categories were combined without weighting them according the Shoemaker routine.[1]

RESULTS

The category scores for students grouped by grade and by reading ability within grade were calculated (for an example, see Tables 5 & 6), as were total item scores, item latencies and individual element scores (see Table 7).

In terms of category scores, the general trend over all categories appears to be one of improvement of item score and decline in latency as subjects progress from grade to grade. This trend is supported by three related findings:

(1) Actual item scores tend to rise as grades rise. For example, 2nd graders score .823 (out of a possible 1.00) on category RI03; 3rd graders score .959; 4th graders score .990. Of the 59 categories for which there are item scores, 39 show a rise from 2nd to 3rd to 4th grade.

(2) Item latencies tend to decline as grades rise. For example, in category RI03, 2nd graders respond in an average of 2.34 seconds, 3rd graders in 1.76 seconds; 4th graders in 1.38 seconds.

(3) Variance in item and latency scores tends to narrow as grades rise.

Where item scores are similar between grades, what separates their performances is the latency score. For example, in category RI02, Grade 3 and 4 item scores are both .918; however, grade 3 mean item latencies are 1.99 (s.d. 1.22) whereas Grade 4's are 1.53 (s.d. .965).

---

[1] For a discussion of this modification, see Walmsley (1975; p. 127).

TABLE 5

Item and latency scores for categories within the CVC universe, by grade (N=48 per grade)

CATEGORY		TOTAL		GRADE 2		GRADE 3		GRADE 4	
(1) REGULAR PRONUNCIATION		Key: I= item score (max = 1.000); L= latency score (in secs)							
REAL WORDS: Frequency as real word		I	L	I	L	I	L	I	L
R101-rare, obsolete, slang	mean	.748	2.29	.734	3.06	.730	2.11	.781	1.71
	s.d.	.251	2.11	.262	3.36	.257	0.89	.229	0.66
R102-low	mean	.890	1.96	.833	2.37	.918	1.99	.918	1.53
	s.d.	.223	1.21	.276	1.47	.185	1.22	.185	0.65
R103-middle	mean	.925	1.82	.823	2.34	.959	1.76	.990	1.38
	s.d.	.214	0.94	.314	1.29	.137	0.62	.071	0.38
R104-high	mean	.945	1.65	.896	1.98	.949	1.64	.990	1.32
	s.d.	.195	0.75	.288	0.99	.151	0.64	.071	0.34
REAL WORDS: Frequency as polysyllable									
R105-doesn't exist	mean	.726	2.38	.667	3.26	.735	2.23	.776	1.68
	s.d.	.446	3.10	.471	4.82	.441	2.00	.417	0.79
R106-exists	mean	.863	2.06	.806	2.65	.823	2.01	.959	1.53
	s.d.	.230	1.14	.271	1.57	.244	0.78	.109	0.47
R107-low	mean	.866	1.88	.833	2.28	.857	1.90	.908	1.56
	s.d.	.276	0.85	.312	0.97	.286	0.83	.218	0.46
R108-middle	mean	.894	1.93	.833	2.51	.888	1.83	.959	1.47
	s.d.	.243	1.14	.295	1.66	.253	0.68	.137	0.38
R109-high	mean	.904	1.81	.854	2.24	.939	1.82	.918	1.39
	s.d.	.294	1.04	.353	1.27	.240	1.05	.274	0.41

TABLE 5 (contd)
Item and latency scores for categories within the CVC universe, by grade

CATEGORY		TOTAL		GRADE 2		GRADE 3		GRADE 4	
(1) REGULAR PRONUNCIATION PATTERN (contd)									
REAL WORDS: Pronunciation Value		I	L	I	L	I	L	I	L
R111-middle	mean	.740	2.57	.729	3.51	.673	2.53	.816	1.69
	s.d.	.439	3.09	.444	3.65	.469	3.64	.387	0.68
R112-high	mean	.877	1.90	.824	2.41	.869	1.82	.936	1.49
	s.d.	.187	0.91	.246	1.19	.170	0.69	.096	0.43
REAL WORDS: Association Value									
R115-middle	mean	.753	2.23	.687	2.81	.735	2.07	.837	1.81
	s.d.	.431	1.12	.464	1.38	.441	0.84	.370	0.81
R116-high	mean	.876	1.99	.818	2.58	.870	1.86	.930	1.53
	s.d.	.175	1.16	.227	1.56	.144	0.77	.116	0.66
NONSENSE SYLLABLES: Association Value									
R117-doesn't exist	mean	.452	3.28	.396	3.88	.449	3.56	.510	2.41
	s.d.	.498	3.05	.489	3.17	.497	3.90	.500	1.26
R118-low	mean	.502	2.99	.500	3.75	.467	2.93	.538	2.31
	s.d.	.256	1.72	.268	1.90	.247	1.76	.248	1.06
R119-middle	mean	.611	2.39	.568	3.07	.620	2.35	.645	1.76
	s.d.	.211	1.50	.210	1.99	.202	1.28	.212	0.59

TABLE 5 (contd)
Item and latency scores for categories within the CVC universe, by grade

CATEGORY (1) REGULAR PRONUNCIATION PATTERN (contd)		TOTAL		GRADE 2		GRADE 3		GRADE 4	
NONSENSE SYLLABLES: Association Value (continued)									
				I	L	I	L	I	L
R120-high	mean	.824	2.04	.792	2.53	.827	2.00	.854	1.62
	s.d.	.205	0.97	.234	1.15	.193	0.84	.180	0.60
NONSENSE SYLLABLES: Pronounceability Value									
R121-low	mean	.288	4.28	.271	5.72	.224	3.70	.367	3.46
	s.d.	.453	4.41	.444	5.66	.417	3.33	.482	3.54
R122-middle	mean	.533	2.84	.537	3.52	.535	2.71	.528	2.30
	s.d.	.220	1.62	.225	1.85	.226	1.66	.209	0.98
R123-high	mean	.654	2.31	.625	2.87	.658	2.28	.679	1.78
	s.d.	.169	1.38	.187	1.44	.183	1.64	.127	0.58
NONSENSE SYLLABLES: Frequency as polysyllable									
R124-doesn't exist	mean	.552	2.79	.535	3.73	.540	2.60	.581	2.05
	s.d.	.206	2.03	.209	2.88	.212	1.42	.193	0.84
R125-exists	mean	.662	2.30	.625	2.90	.656	2.24	.704	1.78
	s.d.	.270	1.34	.286	1.69	.260	1.18	.259	0.75
R126-low	mean	.773	2.41	.708	3.40	.765	2.15	.724	1.70
	s.d.	.347	2.34	.351	3.68	.351	1.12	.336	0.54

TABLE 5 (contd)
Item and latency scores for categories within the CVC universe, by grade

CATEGORY		TOTAL		GRADE 2		GRADE 3		GRADE 4	
(1) REGULAR PRONUNCIATION PATTERN (contd)									
NONSENSE SYLLABLES: Frequency as polysyllable (continued)									
				I	L	I	L	I	L
R127-middle	mean	.788	2.25	.750	3.05	.816	2.12	.796	1.60
	s.d.	.409	2.02	.433	3.05	.387	1.23	.403	0.67
R128-high	mean	.822	2.13	.792	2.80	.796	2.11	.878	1.50
	s.d.	.383	1.82	.406	2.03	.403	2.17	.328	0.56
(2) R-MODIFIED									
REAL WORDS: Frequency as real word									
RM201-4	mean	.767	2.16	.646	3.09	.755	1.87	.898	1.53
	s.d.	.423	1.54	.478	2.24	.430	0.78	.303	0.40
REAL WORDS: Frequency as polysyllable									
RM205-9	mean	.781	2.03	.667	2.66	.755	1.99	.918	1.47
	s.d.	.414	1.53	.471	2.25	.430	1.06	.274	0.47
REAL WORDS: Pronounceability Value									
RM210-12	mean	.856	2.44	.771	4.00	.898	1.88	.898	1.49
	s.d.	.351	6.76	.420	11.57	.303	1.09	.303	0.41
REAL WORDS: Association Value									
RM213-16	mean	.808	2.21	.708	2.82	.816	2.18	.893	1.63
	s.d.	.394	1.63	.455	1.95	.387	1.72	.303	0.73

TABLE 5 (contd)
Item and latency scores for categories within the CVC universe, by grade

CATEGORY		TOTAL		GRADE 2		GRADE 3		GRADE 4	
(2) R-MODIFIED PRONUNCIATION PATTERN (contd)									
NONSENSE SYLLABLES: Association Value		I	L	I	L	I	L	I	L
RM217-20	mean	.760	2.28	.667	3.18	.776	2.09	.837	1.58
	s.d.	.427	1.81	.471	2.57	.417	1.31	,370	0.50
NONSENSE SYLLABLES: Pronounceability Value									
RM221-23	mean	.699	2.49	.500	3.11	.755	2.63	.837	1.74
	s.d.	.459	1.82	.500	1.92	.430	2.19	.370	0.67
NONSENSE SYLLABLES: Frequency as polysyllable									
RM224-8	mean	.459	2.86	.354	3.79	.490	2.83	.531	1.98
	s.d.	.498	2.46	.478	3.30	.500	2.22	.499	0.89
(3) W-MODIFIED									
REAL WORDS: Frequency as real word									
WM301-4	mean	.801	2.15	.687	2.57	.816	2.34	.898	1.55
	s.d.	.399	2.26	.464	1.66	.387	3.34	.303	0.91
REAL WORDS: Frequency as polysyllable									
WM305-9	mean	.555	3.05	.375	4.16	.551	3.30	.735	1.70
	s.d.	.497	4.99	.484	6.30	.497	5.62	.441	0.85
REAL WORDS: Pronounceability Value									
WM310-312	mean	.685	2.49	.562	3.31	.612	3.48	.878	1.70
	s.d.	.465	2.17	.496	2.70	.487	2.27	.328	0.68

TABLE 5 (contd)
Item and latency scores for categories within the CVC universe, by grade

CATEGORY		TOTAL		GRADE 2		GRADE 3		GRADE 4	
(3) W-MODIFIED PRONUNCIATION PATTERN (contd)									
REAL WORDS: Association Value		I	L	I	L	I	L	I	L
WM313-16	mean	.767	2.13	.625	2.62	.816	2.24	.857	1.53
	s.d.	.423	1.88	.484	1.40	.387	2.77	.350	0.60
NONSENSE SYLLABLES: Association Value									
WM317-20	mean	.705	2.66	.562	3.37	.694	2.94	.857	1.70
	s.d.	.456	3.53	.496	2.75	.461	5.26	.350	0.75
NONSENSE SYLLABLES: Pronounceability Value									
WM321-23	mean	.774	2.32	.708	2.99	.714	2.18	.898	1.80
	s.d.	.418	1.59	.455	2.22	.452	1.06	.303	0.94
NONSENSE SYLLABLES: Frequency as polysyllable									
WM324-8	mean	.411	2.41	.313	3.11	.490	2.42	.429	1.73
	s.d.	.492	2.39	.464	2.54	.500	3.06	.495	0.63
(4) Y-MODIFIED									
REAL WORDS: Frequency as a real word									
YM401-4	mean	.801	2.86	.708	3.47	.796	3.41	.898	1.73
	s.d.	.399	4.00	.455	2.80	.403	6.09	.303	0.94
REAL WORDS: Frequency as polysyllable									
YM405-9	mean	.870	2.29	.771	2.84	.898	2.19	.939	1.84
	s.d.	.336	2.64	.420	2.76	.303	2.45	.240	2.62

TABLE 5 (contd)
Item and latency scores for categories within the CVC universe, by grade

CATEGORY		TOTAL		GRADE 2		GRADE 3		GRADE 4	
(4) Y-MODIFIED PRONUNCIATION PATTERN (contd)									
REAL WORDS: Pronounceability Value		I	L	I	L	I	L	I	L
YM410-412	mean	.767	2.33	.729	3.01	.796	2.23	.776	1.76
	s.d.	.423	1.44	.444	1.62	.403	1.45	.417	0.85
REAL WORDS: Association Value									
YM413-16	mean	.904	1.73	.771	2.30	.959	1.61	.980	1.28
	s.d.	.294	1.11	.420	1.64	.198	0.62	.141	0.34
NONSENSE SYLLABLES: Association Value									
YM417-20	mean	.781	2.38	.604	3.25	.857	2.13	.878	1.79
	s.d.	.414	1.48	.489	1.93	.350	0.94	.328	0.93
NONSENSE SYLLABLES: Pronounceability Value									
YM421-23	mean	.616	3.23	.625	4.27	.592	2.68	.633	2.77
	s.d.	.486	3.86	.484	5.81	.491	1.51	.482	2.72
NONSENSE SYLLABLES: Frequency as polysyllable									
YM424-28	mean	.726	2.60	.604	3.46	.714	2.44	.857	1.93
	s.d.	.446	2.29	.489	2.63	.452	2.37	.350	1.43
(5) H-MODIFIED									
REAL WORDS: Frequency as a real word									
HM501-4	mean	.637	2.35	.583	3.23	.551	2.14	.776	1.71
	s.d.	.481	1.46	.493	1.93	.497	1.03	.417	0.68

TABLE 5 (contd)
Item and latency scores for categories within the CVC universe, by grade

CATEGORY		TOTAL		GRADE 2		GRADE 3		GRADE 4	
(5) H-MODIFIED PRONUNCIATION PATTERN (contd)									
REAL WORDS: Frequency as polysyllable		I	L	I	L	I	L	I	L
HM505-9	mean	.664	2.59	.500	3.96	.592	2.18	.898	1.66
	s.d.	.472	3.50	.500	5.57	.491	1.56	.303	0.93
REAL WORDS: Pronounceability Value									
HM510-12	mean	.685	2.49	.604	3.51	.653	2.23	.796	1.76
	s.d.	.465	2.78	.489	4.17	.476	1.55	.403	1.40
REAL WORDS: Association Value									
HM513-16	mean	.664	2.52	.521	3.33	.633	2.44	.837	1.81
	s.d.	.472	1.64	.500	1.78	.482	1.74	.370	0.85
NONSENSE SYLLABLES: Association Value									
HM517-20	mean	.603	2.31	.583	3.05	.551	2.19	.673	1.69
	s.d.	.489	1.64	.493	1.97	.497	1.62	.469	0.81
NONSENSE SYLLABLES: Pronounceability Value									
HM521-23	mean	.596	2.61	.542	3.06	.551	2.91	.694	1.86
	s.d.	.491	3.12	.498	2.02	.497	4.81	.461	0.96
NONSENSE SYLLABLES: Frequency as polysyllable									
HM524-8	mean	.644	2.25	.583	2.85	.551	2.16	.796	1.75
	s.d.	.479	1.40	.493	1.70	.497	1.34	.403	0.77

TABLE 5 (contd)
Item and latency scores for categories within the CVC universe, by grade

CATEGORY		TOTAL		GRADE 2		GRADE 3		GRADE 4	
(6) IRREGULAR PRONUNCIATION									
NONSENSE SYLLABLES: Association Value		I	L	I	L	I	L	I	L
IR618-low	mean	.462	2.94	.433	3.51	.465	2.89	.486	2.42
	s.d.	.297	1.79	.301	1.48	.290	1.79	.297	1.90
IR619-middle	mean	.531	2.92	.500	3.21	.480	3.14	.612	2.41
	s.d.	.406	1.80	.408	1.34	.404	2.28	.395	1.51
NONSENSE SYLLABLES: Pronounceability Value									
IR621-low	mean	.445	2.88	.396	3.63	.429	2.80	.510	2.22
	s.d.	.371	1.44	.367	1.87	.391	1.08	.342	0.78
IR622-middle	mean	.414	3.12	.427	3.84	.418	3.17	.398	2.36
	s.d.	.284	1.89	.284	2.30	.309	1.72	.257	1.19
IR623-high	mean	.589	2.45	.542	3.24	.510	2.22	.714	1.90
	s.d.	.492	1.56	.498	1.93	.500	1.13	.452	1.15
NONSENSE SYLLABLES: Frequency as polysyllable									
IR624-doesn't exist	mean	.504	3.07	.485	4.01	.469	2.96	.557	2.27
	s.d.	.262	1.97	.276	2.34	.269	1.98	.230	0.88

TABLE 6
Item and latency scores for regular pronunciation pattern, grouped by reading ability within grade

		CATEGORY (1) REGULAR PRONUNCIATION REAL WORDS: Frequency as real word							
		R101- rare, obs, slang		R102- low		R103- middle		R104- high	
GOOD READERS									
-Grade 2	mean	.839	1.74	.929	1.42	.929	1.42	1.00	1.33
N=8	s.d.	.203	0.40	.175	0.23	.175	0.20	0.0	0.17
-Grade 3	mean	.821	1.81	.929	1.49	.964	1.47	1.00	1.33
N=16	s.d.	.175	0.83	.175	0.80	.129	0.25	0.0	0.32
-Grade 4	mean	.842	1.51	.947	1.36	1.00	1.23	1.00	1.22
N=8	s.d.	.185	0.36	.153	0.34	0.0	0.21	0.0	0.22
AVERAGE READERS									
-Grade 2	mean	.773	2.42	.886	2.11	.841	2.08	1.00	1.70
N=22	s.d.	.237	0.82	.210	0.67	.277	0.67	0.0	0.40
-Grade 3	mean	.824	1.84	.971	1.64	.971	1.67	.941	1.49
N=17	s.d.	.187	0.34	.118	0.39	.118	0.33	.161	0.32
-Grade 4	mean	.775	1.87	1.00	1.54	.975	1.47	.975	1.45
N=20	s.d.	.236	0.66	0.0	0.53	.109	0.42	.109	0.43
POOR READERS									
-Grade 2	mean	.406	6.20	.750	3.69	.750	3.79	.563	3.19
N=14	s.d.	.174	7.01	.250	1.53	.354	1.21	.464	1.00
-Grade 3	mean	.562	2.50	.844	2.24	.969	1.92	.906	1.96
N=14	s.d.	.300	0.97	.232	0.95	.121	0.73	.195	0.79
-Grade 4	mean	.688	1.73	.750	1.81	1.00	1.54	1.00	1.28
N=19	s.d.	.272	1.00	.250	1.19	0.0	0.51	0.0	0.26

TABLE 6 (contd)
Item and latency scores for regular pronunciation pattern, grouped by reading ability within grade

CATEGORY REAL WORDS: Frequency as polysyllable

		R105 does not exist		R106 exists		R107 low		R108 middle	
GOOD READERS									
-Grade 2	mean	.857	1.61	.952	1.54	.893	1.53	.964	1.46
	s.d.	.350	0.37	.117	0.29	.279	0.23	.129	0.21
-Grade 3	mean	.929	1.61	.881	1.65	1.00	1.56	.964	1.55
	s.d.	.258	0.42	.160	0.56	0.0	0.32	.129	0.40
-Grade 4	mean	.895	1.53	.965	1.37	.921	1.33	.974	1.40
	s.d.	.307	0.59	.102	0.24	.244	0.29	.112	0.41
AVERAGE READERS									
-Grade 2	mean	.727	2.18	.879	2.20	.909	2.17	.818	2.23
	s.d.	.445	0.91	.160	0.56	.193	0.74	.321	1.28
-Grade 3	mean	.706	1.84	.824	1.81	.765	1.72	.941	1.67
	s.d.	.456	0.46	.259	0.40	.348	0.36	.161	0.42
-Grade 4	mean	.750	1.63	.967	1.58	.925	1.53	.975	1.53
	s.d.	.433	0.46	.100	0.42	.179	0.47	.109	0.33
POOR READERS									
-Grade 2	mean	.250	8.25	.542	4.61	.688	3.16	.750	4.03
	s.d.	.433	10.05	.286	1.77	.348	0.42	.250	1.20
-Grade 3	mean	.625	2.97	.771	2.46	.812	2.12	.750	2.07
	s.d.	.484	3.21	.282	0.99	.300	0.89	.354	0.75
-Grade 4	mean	.625	2.03	.958	1.60	.875	1.60	.875	1.45
	s.d.	.484	1.47	.110	0.71	.217	0.66	.217	0.40

TABLE 6 (contd)
Item and latency scores for regular pronunciation pattern, grouped by reading ability within grade

		CATEGORY REAL WORDS:					
		Frequency as polysyllable		Pronounceability Value			
		R109 high		R111 middle		R112 high	
GOOD READERS							
-Grade 2	mean	.929	1.45	.857	1.67	.908	1.54
	s.d.	.258	0.27	.350	0.63	.139	0.30
-Grade 3	mean	1.00	1.55	1.00	1.79	.918	1.61
	s.d.	0.0	0.60	0.0	0.69	.159	0.72
-Grade 4	mean	1.00	1.25	.842	1.47	.955	1.33
	s.d.	0.0	0.23	.365	0.40	.081	0.29
AVERAGE READERS							
-Grade 2	mean	.864	2.17	.864	2.86	.883	2.06
	s.d.	.343	1.30	.343	1.84	.165	0.56
-Grade 3	mean	1.00	1.52	.647	1.82	.891	1.63
	s.d.	0.0	0.22	.478	0.63	.125	0.27
-Grade 4	mean	.900	1.43	.800	1.82	.936	1.57
	s.d.	.300	0.42	.400	0.65	.071	0.46
POOR READERS							
-Grade 2	mean	.750	2.80	.500	8.19	.643	3.91
	s.d.	.433	0.64	.500	6.14	.267	0.64
-Grade 3	mean	.812	1.97	.375	3.83	.830	1.97
	s.d.	.390	0.86	.484	6.07	.184	0.62
-Grade 4	mean	.750	1.58	.750	1.95	.929	1.53
	s.d.	.433	0.58	.433	1.07	.101	0.51

TABLE 6 (contd)
Item and latency scores for regular pronunciation pattern, grouped by reading ability within grade

		CATEGORY REAL WORDS: Association Value				NONSENSE SYLLABLES: Association Value			
		R115 middle		R116 high		R117 doesn't exist		R118 low	
GOOD READERS									
-Grade 2	mean	.643	1.84	.878	1.59	.286	2.99	.589	2.50
	s.d.	.479	0.49	.108	0.28	.499	1.89	.257	0.76
-Grade 3	mean	.786	1.97	.949	1.50	.357	2.99	.607	2.44
	s.d.	.410	0.89	.068	0.35	.479	3.17	.236	1.13
-Grade 4	mean	.947	1.76	.962	1.35	.474	2.16	.638	1.80
	s.d.	.233	0.65	.063	0.28	.499	1.19	.243	0.51
AVERAGE READERS									
-Grade 2	mean	.818	2.50	.883	2.20	.500	3.49	.528	3.52
	s.d.	.386	1.10	.153	1.13	.500	3.06	.226	1.56
-Grade 3	mean	.765	1.69	.874	1.62	.529	2.52	.449	2.36
	s.d.	.424	0.43	.108	0.20	.499	1.29	.257	0.49
-Grade 4	mean	.900	1.77	.943	1.52	.500	2.56	.462	2.51
	s.d.	.300	0.60	.095	0.31	.500	1.21	.217	1.14
POOR READERS									
-Grade 2	mean	.625	4.54	.679	4.08	.250	4.85	.328	5.66
	s.d.	.484	0.81	.255	1.33	.433	2.24	.176	1.71
-Grade 3	mean	.687	2.37	.804	2.12	.375	4.49	.367	3.15
	s.d.	.464	0.86	.188	0.60	.484	5.36	.200	1.38
-Grade 4	mean	.375	1.98	.857	1.91	.625	2.38	.469	2.55
	s.d.	.484	1.43	.160	1.40	.484	1.23	.263	1.07

TABLE 6 (contd)
Item and latency scores for regular pronunciation pattern, grouped by reading ability within grade

		CATEGORY NONSENSE SYLLABLES							
		Association Value				Pronounceability Value			
		R119 middle		R120 high		R121 low		R122 middle	
GOOD READERS									
-Grade 2	mean	.670	1.94	.857	1.70	.286	4.10	.659	2.29
	s.d.	.230	0.40	.198	0.28	.452	2.95	.212	0.49
-Grade 3	mean	.723	1.78	.869	1.59	.214	3.04	.627	2.09
	s.d.	.178	0.54	.157	0.37	.410	2.25	.243	0.64
-Grade 4	mean	.697	1.52	.868	1.38	.474	2.15	.596	1.84
	s.d.	.204	0.32	.176	0.32	.499	0.69	.185	0.45
AVERAGE READERS									
-Grade 2	mean	.568	2.57	.848	2.18	.318	5.70	.530	3.24
	s.d.	.187	1.17	.150	0.61	.466	6.37	.167	1.46
-Grade 3	mean	.632	2.15	.863	1.78	.294	2.84	.549	2.31
	s.d.	.184	0.77	.143	0.48	.456	1.00	.233	0.43
-Grade 4	mean	.644	1.89	.850	1.69	.350	4.98	.522	2.61
	s.d.	.174	0.46	.174	0.50	.477	4.98	.225	1.04
POOR READERS									
-Grade 2	mean	.438	5.58	.625	4.08	.125	7.55	.417	5.63
	s.d.	.140	2.83	.217	0.72	.331	6.81	.227	2.14
-Grade 3	mean	.516	2.61	.729	2.27	.187	4.67	.444	3.32
	s.d.	.182	0.90	.235	0.74	.390	4.90	.176	2.39
-Grade 4	mean	.563	1.88	.833	1.80	.250	2.23	.431	2.64
	s.d.	.280	1.06	.204	1.01	.433	0.70	.161	1.33

TABLE 6 (contd)
Item and latency scores for regular pronunciation pattern, grouped by reading ability within grade

CATEGORY NONSENSE SYLLABLES

		Pronounceability Value		Frequency as polysyllable					
		R123 high		R124 doesn't exist		R125 exists		R126 low	
GOOD READERS									
-Grade 2	mean	.696	1.90	.626	2.20	.821	1.78	.643	1.78
	s.d.	.162	0.32	.211	0.59	.133	0.40	.350	0.55
-Grade 3	mean	.756	1.69	.637	2.03	.810	1.76	.786	1.70
	s.d.	.177	0.40	.239	0.60	.177	0.45	.311	0.70
-Grade 4	mean	.741	1.52	.644	1.80	.702	1.51	.842	1.52
	s.d.	.085	0.34	.194	0.70	.227	0.38	.232	0.40
AVERAGE READERS									
-Grade 2	mean	.655	2.55	.566	3.20	.561	2.63	.841	2.56
	s.d.	.165	0.97	.170	1.39	.287	1.62	.277	1.10
-Grade 3	mean	.672	1.89	.529	2.20	.598	2.04	.794	1.94
	s.d.	.160	0.35	.206	0.51	.275	0.90	.300	0.55
-Grade 4	mean	.671	1.89	.562	2.12	.692	2.01	.750	1.73
	s.d.	.125	0.56	.189	0.88	.313	0.87	.335	0.51
POOR READERS									
-Grade 2	mean	.510	4.39	.404	7.04	.542	4.28	.500	6.24
	s.d.	.141	1.50	.175	4.74	.232	0.95	.354	5.15
-Grade 3	mean	.568	2.47	.481	2.91	.625	2.50	.617	2.29
	s.d.	.172	0.72	.157	1.19	.239	0.86	.428	0.58
-Grade 4	mean	.583	1.93	.510	2.27	.750	1.88	.438	1.88
	s.d.	.138	0.75	.172	0.88	.204	0.93	.390	0.69

TABLE 6 (contd)
Item and latency scores for regular pronunciation pattern, grouped by reading ability within grade

		CATEGORY	NONSENSE SYLLABLES			
			Frequency as polysyllable			
			R127 middle		R128 high	
GOOD READERS						
-Grade 2	mean	1.00	1.65		.857	1.61
	s.d.	0.0	0.27		.350	0.38
-Grade 3	mean	.786	1.96		.786	2.52
	s.d.	.410	1.18		.410	3.90
-Grade 4	mean	.895	1.55		.842	1.33
	s.d.	.307	0.72		.365	0.32
AVERAGE READERS						
-Grade 2	mean	.727	2.45		.773	2.36
	s.d.	.455	1.50		.419	1.08
-Grade 3	mean	.941	1.94		.824	1.59
	s.d.	.235	1.00		.381	0.31
-Grade 4	mean	.800	1.52		.950	1.61
	s.d.	.400	0.57		.218	0.58
POOR READERS						
-Grade 2	mean	.375	4.55		.750	5.00
	s.d.	.484	2.89		.433	3.17
-Grade 3	mean	.750	2.16		.750	2.18
	s.d.	.433	0.75		.433	0.59
-Grade 4	mean	.625	1.66		.875	1.61
	s.d.	.484	0.58		.331	0.85

When the results are inspected from the perspective of subjects grouped by reading ability within grade, a similar picture emerges. There are, however, some additional findings when the subjects are grouped this way:

(1) The mean item scores of good readers from Grades 2, 3 and 4 are not markedly different from one another. For example, in category RIOI, Grade 2's good readers score .839; Grade 3's, .821, and Grade 4's .842.

(2) The mean item latencies of good readers from Grades 2, 3 and 4 are not markedly different from one another. For example, in category RIOI, Grade 2's good readers have a mean item latency of 1.74 secs' Grade 3's have 1.81; and Grade 4's have 1.51.

(3) The mean item scores of average readers from grades 2, 3 and 4 are not markedly different from one another, although the mean item latencies of 2nd graders are consistently slower than those of average readers in grades 3 and 4.

(4) The mean item and latency scores for poor readers from the three gardes are widely separated from each other. While the mean item scores of 3rd and 4th grade poor readers are not always higher than those of poor readers (e.g. in category RII5, poor readers in Grade 2 score .625, poor readers in grade 4 score .375), latencies show a consistent trend. 4th grade poor readers always respond to items more quickly than poor readers in the 2nd and 3rd grades.

Total scores

Table 7 presents the means and standard deviations of the total latency scores and totals for the individual CVC elements (initial consonants, medial vowels, final consonants), grouped by grade and by reading ability within grade.

To summarize these results; it appears that by the middle of the academic year in Grade 2, the average 2nd grader in the sample has a better than 60% understanding of the pronunciations of CVCs. By the middle of the 3rd grade academic year, this understanding has improved to better than 65% and by the middle of the 4th grade it has further improved to better than 70%. The overall difference between the three

TABLE 7

Comparison of total item scores, latency scores and individual element scores, by grade and reading ability within grade

SCORE		COMBINED 2nd, 3rd & 4th Grades	Grade 2	Grade 3	Grade 4	READING ABILITY WITHIN GRADE								
						-----GRADE 2-----			-----GRADE 3-----			-----GRADE 4-----		
						POOR	AVGE	GOOD	POOR	AVGE	GOOD	POOR	AVGE	GOOD
TOTAL SCORE	mean	.666	.623	.659	.715	.443	.661	.734	.556	.683	.766	.638	.708	.769
	s.d.	.158	.183	.152	.118	.138	.122	.124	.144	.121	.112	.120	.105	.099
TOTAL LATENCY	mean	2.48	3.18	2.40	1.88	5.26	2.81	1.98	2.75	2.03	1.88	2.04	2.01	1.60
	s.d.	1.36	1.66	1.24	0.63	1.63	0.96	0.37	1.10	0.39	0.52	0.91	0.64	0.34
TOTAL INITIAL CONSONANT	mean	.912	.903	.905	.928	.862	.911	.934	.860	.925	.943	.903	.924	.944
	s.d.	.054	.056	.064	.034	.047	.033	.053	.073	.037	.018	.022	.039	.023
TOTAL MEDIAL VOWEL	mean	.743	.700	.746	.782	.549	.730	.798	.664	.760	.824	.735	.779	.818
	s.d.	.135	.164	.124	.098	.132	.107	.112	.119	.103	.103	.097	.085	.092
TOTAL FINAL CONSONANT	mean	.922	.913	.910	.942	.859	.926	.950	.856	.930	.956	.910	.940	.959
	s.d.	.073	.082	.080	.049	.107	.064	.039	.100	.046	.043	.050	.045	.046

grades is significant ($p = < .015$); posthoc comparisons reveal that only the difference between Grade 2 and Grade 4 is significant ($.01 < p < .05$).[1]

At the same time, the speed with which a subject recognizes and is able to pronounce a CVC has dropped from an average of over 3 seconds in Grade 2, to an average of under 2 1/2 seconds in Grade 3, and finally to an average of under 2 seconds in Grade 4 (Overall significance, $p = < .001$; posthoc comparisons--Grade 2 latencies are significantly longer than Grade 3 or Grade 4, $p = < .01$; Grade 3's are not significantly longer than Grade 4's).

CVC element scores (initial consonant, medial vowel, final consonant) are relatively constant across grades, but there are differences in element scores within grade and within reading ability in grades. Most errors are made on medial vowels (the mean medial vowel score for the combined grades is .743). There is a significant difference between the overall medial vowel scores of the three grades ($p = < .012$); however, the posthoc comparisons reveal that only grade 2 and grade 4 medial vowel scores are significantly different ($p = < .01$).

Scores on the initial consonant are very close to those on the final consonant; the combined grade score .912 on initial consonants, .922 on final consonants. Differences between the three grades in initial consonant scores are significant overall ($p = < .039$), but posthoc comparisons reveal no significant differences between the individual grades. Differences between grades in the final consonant scores were not significant.

[1] The differences in means were tested for statistical significance by the one-way analyses of variance. Newman-Keuls pairwise posthoc comparisons (Kirk, 1968; p. 91) were carried out to determine the significance of differences between individual grade performances.

When the results are analyzed according to reading ability, a somewhat sharper picture emerges. There are significant differences overall, between good, average and poor readers on total score ($p < .0001$); on total latencies ($p < .0001$); on medial vowel scores ($p < .0001$); on initial consonant scores ($p < .001$) and on final consonant scores ($p < .006$).

Within the 2nd and 3rd grade, there are significant differences between good, average and poor readers on total score, total latency, and the individual element scores.

In the 4th grade, there are barely significant differences between good, average and poor readers on total scores ($p < .0377$); however, no other differences reached significance.

DISCUSSION

The discussion is divided into sections which deal with:

(a) the definition of the CVC universe and its categoreis;

(b) test construction procedures and their difficulties;

(c) interpretation of the results, from both reading and statistical perspectives;

(d) procedures appropriate for the development of a final form of a criterion-referenced measure of the CVC universe.

(a) Definition of the CVC universe and its categories

One of the purposes of administering the CVC test to a representative sample of students was to determine whether the dimensions of the CVC universe as I had construed them were meaningful subcategories of the CVC universe. If they were, then a final test of this universe

should make provision for them; if not, then either other dimensions should be sought, or it must be concluded that the CVC universe is a unified entity from which a random sample of items may be drawn to represent it adequately.

The test results presented in the previous section indicate that there are some differences (some statistically significant, some not) between subjects, whether grouped by grade or by reading ability within grade. These results also indicate that performances differ between categories of the CVC universe. A strong trend suggests itself from an examination of the data: namely, that categories in which CVCs are grouped according to some measure of their "meaningfulness" appear to yield higher category scores, the more meaningful they are.

One way of defining "meaningfulness" is whether a CVC is a real word or a nonsense syllable. Defined this way, there are differences between the real word categories and the nonsense syllable categories. These results are displayed in Table 8.

Another way of defining meaningfulness is whether a CVC has high or low association value (i.e. whether a subject is "reminded" of a real word when he sees a nonsense syllable), or high or low frequency as a syllable in a longer word. It will be seen from Table 8 that categories which are high association value or high frequency as a polysyllable (whether real words or nonsense syllables) yield higher mean item scores than those with low association values or low frequency as a polysyllable.

It is clear that the dimension "meaningfulness" (however it is defined) is an important factor in determining whether a child can pronounce a CVC successfully, and must therefore be regarded as an indication that the CVC universe needs to be broken into categories

TABLE 8
Category item scores ranked by grade and reading ability within grade

CATEGORY	Rank	Combined 2nd, 3rd 4th grade	Range	Grade 2	3	4	Range	Reading Ability within Grade High	Average	Low
REAL WORDS										
R104 (high frequency as real word)	1	.945	(.748-.945)	.896	.949	.990	(.552-1.00)	1.00	.972	.906
R109 (high frequency as polysyllable)	2	.904	(.726-.904)	.854	.939	.918	(.500-.976)	.976	.921	.791
R112 (high pronounceability value)	3	.877	(.740-.877)	.824	.869	.936	(.542-.927)	.927	.903	.801
R116 (high association value)	4.	.876	(.753-.876)	.818	.878	.930	(.562-.930)	.930	.900	.780
NONSENSE SYLLABLES										
R120 (high association value)	5.	.824	(.452-.824)	.792	.827	.854	(.372-.865)	.865	.854	.729
R128 (high frequency as polysyllable)	6.	.822	(.552-.822)	.792	.796	.878	(.465-.894)	.894	.849	.792
R123 (high pronounceability value)	7.	.654	(.288-.654)	.625	.658	.679	(.187-.731)	.731	.666	.554

which reflect this dimension. However, the problem remains: which definition of meaningfulness do we adopt?

In terms of the utility of dividing the CVC universe into its various pronunciation patterns, a case has already been made earlier for distinguishing between the checked vowel pronunciation pattern and modified pronunciation patterns. In other words, CVCs such as bat clearly represent different pronunciation patterns from those such as bar, baw, or bay. These distinctions need to be made regardless of the performances of examinees on a CVC test which includes all these various patterns.

However, some of the test's findings indicate that these modified pronunciation patterns are not as easy to pronounce as the checked vowel pronunciation pattern. Performances on categories where equivalencies are possible to observe are generally lower for the modified vowel categories than for the checked vowel. This only seems contraindicated by the modified-y pronunciation pattern.

Although meaningful distinctions can be made between the various pronunciation patterns, there are insufficient items within the h-modified, w-modified and y-modified patterns to support those distinctions empirically (i.e. to say whether students find one pattern more difficult to pronounce than another, holding the dimension of "meaningfulness" constant.) In addition, the ambiguity of w and y as final consonants makes them poor candidates for a final CVC test.

R-modified CVCs appears to be better suited for such a test; the r-modified CVC item pool is reasonably sized, and presents less ambiguity than w or y with respect to its final consonant sound.

This leaves the irregular pronounciation pattern; it is a very sizeable pool (456 items), which seems (appropriately) to contain the

CVCs which are hardest to pronounce. While the results generally confirm the appropriateness of the criteria for placing CVCs into this category, this item pool may usefully be discarded as far as the development of a final form of the CVC test is concerned.

In summary, we must conclude from this discussion that the division of CVC items into various pronunciation patterns, and into categories reflecting their meaningfulness are both important first steps in the construction of a criterion-referenced test of the CVC universe. At the same time, it would appear that some collapsing of categories, and perhaps the omission of certain categories (i.e. irregular, h-modified) might be in order. Procedures which might assist in this task are discussed later.

(b) Test construction procedures and their difficulties

1. Item generation procedures

The generation of all possible items within the CVC universe yielded 2,646 items, each of which had to be classified into 19 categories, many of which had three divisions (high, medium, low), making a total of 59 categories. The procedure was carried out as precisely as possible, to show how a criterion measure should be constructed.

Several difficulties arose. Firstly, many decisions had to be made on questions of inclusion/exclusion in categories, beginning with defining what constituted a consonant in a CVC, and ending by examining every CVC in Webster's dictionary to determine whether it existed as a word. Secondly, the procedure results in an enormous number of items in very many categories. To state the problem another way; the universe chosen was comparatively small compared with other

universes in reading--for example, the addition of just one more letter (CCVC, or CVCC) would have yielded an item pool in excess of 12,000. The stratified random sample adopted in this study does, however, represent a good compromise between a truly random sample and an arbitrary selection of items. Nevertheless, there are no shortcuts available in the task of determining the categories within which the various item pools are nested. In summary, the CVC universe as defined in this study is not amenable to simple random sampling, although it would be interesting to speculate whether a simple random sample of the CVC universe would give results similar to those found empirically here. While the overall scores might be similar, the results might be very misleading with respect to various categories subsumed within the universe. The stratified random sampling plan may be a lengthy procedure but it at least ensures that different types of CVC are (a) sampled separately, and (b) represented according to their relative weight in the total universe.

2. Item analysis procedures

The present study was not able to contribute much to the discussion of the appropriateness of various item analysis techniques for criterion-referenced measurement (i.e. factor analysis; Rasch's sample-free test calibration), primarily because the experimental measure was not in its final form. Consequently, determinations of item homogeneity and item-objective congruence were not appropriate at this stage.

However, the differences between norm-referenced and criterion-referenced item analysis procedures is clearly demonstrated later

In the section on procedures for developing a final form of the test.

(c) Interpretation of results from a reading perspective

Although not a major focus of the study, there were several questions raised by the construction of a criterion measure of the CVC universe, to which the study might at least offer some tentative and speculative answers.

The data presented in the Results section would suggest that an understanding of the pronunciation of CVCs improves with age and grade placement: even poor readers in the sample managed by the middle of 4th grade to have acquired a reasonable competence with the pronunciation of CVCs. However, the performances of good readers in the 2nd, 3rd and 4th grades are very much alike--the good readers in grade 2 being very close to good readers in grade 4 in almost every respect, including latencies.

That more vowel errors occurred than initial or final consonant errors is to be expected; these findings confirm what many others have reported earlier (e.g. Liberman, 1973, p.73). However, that almost equal numbers of errors are made by subjects on initial and final consonants is a surprising result, which conflicts with the reported findings of others (Weber, 1970; Liberman, 1973) who suggest that in oral reading, final consonants are more frequently missed than initial consonants. My data show a trend that might suggest the opposite is true.

The inconsistencies between my findings and those of others may lie in one of the following differences between this study and theirs:

(1) The studies which report higher final consonant error rates used real words, either in the context of connected prose (Weber, 1970), or in lists of words familiar to subjects (Liberman, 1973). My study used a combination of real words and nonsense syllables.

(2) My items were drawn randomly from categories without controlling for the frequency of specific initial or final consonants. It may be that my items are not normally distributed in this respect.

A second issue relates to the category scores of subjects in my sample. In only one or two categories out of the 59 were perfect scores obtained, whether subjects were grouped by grade or by reading ability within grade. This would imply that school children similar to those in my sample do not master CVCs from various categories (except highly frequent real words) with 100% accuracy. Therefore, to set mastery levels that are higher than the typical performance of the best readers would seem to require some justification.

Interpretation of category scores

The findings relating to the category scores suggest that the dimensions of the CVC universe as proposed earlier were indeed viable categories, but they also suggest:

(1) That if CVCs are grouped according to the major categories, and if school children perform differently on these categories, then tests which are devised to predict one category score from another may only do so if the categories are equivalent, or related to one another. In other words, if CVCs that are real words are indeed different from CVCs that are nonsense syllables, then caution needs to be exercised in testing children on nonsense syllables and generalizing the results to their ability to read real words;

(2) that children sometimes pronounce a CVC on the basis of

its approximation to a real word (or perhaps a highly frequent word that does not follow the generalization) rather than on the basis of the phonics generalization. For example, the CVC ced, which should be pronounced /sed/ was more often pronounced /ked/ by subjects in my sample. The implication here is that the child is relying on a different set of principles than the ones provided by the phonics generalization. In other words, the child may be pronouncing ced on the basis of the letter c always being pronounced /k/ when it appears in three letter words (i.e. cap, cup, cot). The same appears to hold true with the vowel y which in the medial position was almost invariably pronounced long /iy/, probably on the basis of its overall frequency with this pronunciation (i.e. in words such as try, shy, etc.).

(3) that some initial consonants (e.g. x) give children so much difficulty that their knowledge of the checked vowel which follows the initial x may be obscured by their inability to cope with the initial consonant.

These observations suggest that many of the items which were so carefully drawn from categories, did indeed belong to categories-within-categories (or at least were so perceived by subjects).

One important implication here is that before testing a child on nonsense syllables (or any non-real word category), some determination of the relationship between nonsense syllables and real words, and the hierarchies between categories needs to be made. One such procedure for estimating hierarchies is Bart's ordering theory (Bart & Krus, 1973). This technique analyses the category scores and determines the logical relationship between them. In essence, it examines the proportion of subjects who scored one category correct and another incorrect; these proportions are termed disconfirmatory responses, and from a table of these response patterns, an estimate of the hierarchical nature of the categories may be determined.

In this case, a mastery level of 75% was set, and the appropriate statistics computed. These are displayed in Tables 9 and 10.

TABLE 9

Matrix of percentages of disconfirmatory reponse patterns for prerequisite relation orderings for categories

CATEGORIES			101	102	103	104	105	106	107	108	109	111	112	115	116	117	118	119	120	121	122	123	124	125	126	127	128
REAL WORDS	Frequency as real word	-rare	101		15	23	23	16	16	19	17	23	17	17	16	14	9	5	4	16	3	3	3	3	10	10	21	21
		-low	102	7		14	16	11	10	14	11	16	12	10	11	10	7	3	1	10	1	2	1	1	3	6	14	14
		-middle	103	6	6		10	8	3	8	5	9	7	4	10	6	4	1	1	8	1	0	1	0	1	4	8	8
		-high	104	2	3	6		1	2	3	4	7	3	3	5	3	3	0	1	3	0	1	1	0	1	4	3	6
	Frequency as a polysyllable	-dsnt exist	105	15	18	23	20		15	21	19	23	14	19	16	16	9	4	3	14	4	1	6	1	10	14	16	21
		-exists	106	18	20	22	25	19		23	21	25	17	19	20	19	9	1	6	17	2	2	5	1	8	15	22	21
		-low	107	12	14	16	16	14	13		12	18	14	13	14	12	8	3	3	10	2	1	3	2	5	8	14	16
		-middle	108	6	8	10	14	10	8	8		12	7	6	11	6	4	1	1	7	0	0	1	0	3	4	11	12
		-high	109	3	5	6	8	5	4	6	4		5	5	5	3	2	1	1	4	1	0	1	0	3	1	6	6
	Pron Value	-middle	111	14	17	21	21	13	12	19	15	21		15	16	14	10	3	4	14	3	4	6	1	8	12	20	17
		-high	112	8	9	11	14	10	8	11	8	14	8		12	8	7	2	2	10	1	0	3	0	3	6	12	12
	Assoc Value	-middle	115	12	15	22	21	13	14	17	18	20	14	17		16	6	4	7	14	4	3	8	5	11	8	16	17
		-high	116	6	10	14	15	10	9	12	9	14	9	9	12		4	2	1	10	1	1	1	0	4	6	14	13
NONSENSE SYLLABLES	Assoc Value	-dsnt exist	117	35	41	37	50	36	43	41	41	47	38	43	36	38		8	13	33	8	8	17	8	27	25	45	45
		-low	118	50	56	63	96	51	45	56	57	65	51	57	53	56	27		13	50	11	8	20	8	25	43	54	56
		-middle	119	41	47	55	58	42	40	47	49	57	44	49	48	47	24	5		42	11	6	15	6	19	38	47	50
		-high	120	14	15	21	21	12	12	51	15	21	14	17	15	16	4	2	2		3	1	3	1	6	8	16	17
	Pron a polysyllable Value	-low	121	46	52	60	63	48	43	52	53	62	49	53	51	52	25	8	16	48		10	19	9	27	39	56	56
		-middle	122	53	61	67	72	53	51	59	62	70	58	60	58	60	33	13	19	54	19		22	10	31	44	60	62
		-high	123	38	45	53	57	43	38	47	47	56	45	48	48	45	27	10	14	42	12	7		1	21	36	51	50
	Frequency as a polysyllable	-dsnt exist	124	53	60	67	71	53	49	60	62	70	55	60	60	59	33	13	19	55	17	10	16		30	47	62	63
		-exists	125	35	37	63	47	37	31	38	40	47	36	38	40	38	26	6	8	34	10	6	10	4		33	39	40
		-low	126	23	27	34	38	28	26	28	28	34	27	29	25	27	12	11	14	24	10	6	23	9	21		31	31
		-middle	127	13	14	17	16	10	12	14	14	18	15	14	12	15	11	1	3	12	6	1	9	4	6	10		15
		-high	128	10	11	13	15	11	8	12	12	14	9	10	10	10	8	1	2	9	3	1	3	1	3	7	12	

TABLE 10

Summary of prerequisite relationships between categories in CVC universe

CATEGORY	IS PREREQUISITE TO:
R101	-
R102	R119 R121 R123 R124
R103	R118 R119 R121 R122 R123 R124 R125
R104	R105 R118 R119 R121 R122 R123 R124 R125
R105	R122 R124
R106	R122
R107	R122
R108	R118 R119 R121 R122 R123 R124
R109	R118 R119 R121 R122 R123 R124 R126
R111	R124
R112	R121 R122 R124
R115	-
R116	R119 R121 R122 R123 R124
R117	-
R118	-
R119	-
R120	R122 R124
R121	-
R122	-
R123	R124
R124	-
R125	-
R126	-
R127	R118 R122
R128	R118 R122 R124

Tolerance= .05

It will be seen from these tables that several prerequisite relationships exist between categories of the CVC universe. There appears to be an overall trend that real word categories appear to be consistently prerequisite to nonsense syllable categories. These relationships imply that my subjects had to have scored well on real word categories if they were to score well on nonsense syllable categories. In other words, the two types of category are not independent of one another: real word CVCs are easier to pronounce than nonsense syllables.

Interpretation of the results from a statistical perspective

Although a number of issues relating to the construction of criterion-referenced tests were not addressed by this study, there are some conclusions to be drawn from the exercise. Firstly, the technique of stratified random sampling of universes seems eminently suitable for the type of measurement required in the present study. Although it resulted in a very large test (914 items) and large tests have built-in difficulties, the application of this procedure on a large scale (e.g. the development of a test to be given to large numbers of students) seems not to present overwhelming difficulties. In addition, there is no reason why such a test should be too large, provided that the procedure employed to reduce its size does not interfere with the principle of items adequately representing universes.

Secondly, the discussion earlier on the question of validity and especially on the relationship between items and categories, only underscores the importance of the development of item analysis

procedures which estimate not so much the correlations between item scores and total scores but rather whether all the items in a category measure the same thing. If they do, then a category may be said to be truly unidimensional. However, as has been noted, subjects do not always respond to the same categories as the test constructors had hypothesized they would; consequently, it will always be extremely difficult to measure traits, especially ones so minute as those measured here, uncontaminated by other traits. It may also be observed that the more minutely a universe is fractionated, the greater the variety of errors is possible, for elements are more susceptible to be influenced by the environments in which they are placed. These raise some difficult questions for the development of valid and reliable tests in early reading, and suggest that we may need to interpret results of criterion measures with much more sensitivity to and tolerance for errors of measurement than we are expecting we should. It is paradoxical that tests which are intended to be more precise appear to produce results which are so highly susceptible to error. However, the experience with CVCs leads me to speculate that this might indeed be the case.

(d) Procedures appropriate to the development of a final form of a CVC test

Several suggestions for developing a final form of the test are prompted by the study.

Firstly, an examination of the category scores has revealed that although there are major differences between categories which contain real words, and those appear to be substantially different from one another (i.e. association value, pronounceability value). Thus the

first step in deciding whether any categories should be collapsed or dropped, would be to examine the relative strength of a category in the hierarchical order of difficulty, as determined through ordering theory. Some experimentation with mastery levels and different groupings of subjects will be necessary here.

A second step would be to examine the actual responses given to items. This examination would focus on two indices--item difficulty and item discriminating power. Here, of course, one will have to make a decision as to whether the final test is intended to be norm-referenced or criterion-referenced. Up to this point, the procedure for developing either kind of test have been appropriate (althought unnecessarily laborious and elaborate for NRM). However, at this stage, one may select those items which maximally discriminate between student performances (i.e. are norm-referenced), and those which best represent the universe (i.e. are criterion-referenced). Items with medium difficluty and high discriminating power will be appropriate for the former; items which do not show errors due to anything (as best as may be determined) but a lack of knowledge of the particular category will be appropriate to the latter. That these two procedures produce different items may be more clearly demonstrated by providing an example, based on item difficulties and discriminating powers of all items in category RIOI. These will be found in Table II

There are threats associated with both types of item selection. The norm-referenced procedure will, of course, result in a test with good predictive and selective powers, but it may not necessarily represent each category adequately, since the randomization process has been eliminated. Also, as may be seen in Table II, the items

TABLE II

Item difficulties and discriminating powers for all items in category R101 (Regular Pronunciation Pattern: Real Words by Frequency Value, rare, obs)

TRAY	ITEM #	ITEM	DIFF	CAT DISCR	TOT DISCR	CRTEST	NRTEST	REMARKS
1	113	HOD	.808	.735	.478		1	
2	113	VON	.917	.529	.479			
3	113	TAJ	.609	.743	.493		1	
4	113	CIT	.292	.524	.151			29% init con err
5	113	GUV	.792	.497	.235	1		
6	113	MAM	.917	.787	.415	1		
1	119	TYG	.115	.321	.093			57% midvowel long
2	119	COQ	.542	.755	.447		1	25% last con err
3	119	TAV	.696	.727	.447	1	1	
4	119	VAC	.875	.777	.676		1	
5	119	KOB	.750	.712	.577	1	1	
6	119	JON	.875	.461	.358	1		
1	90	TUP	.923	.201	.238	1		
2	90	GAN	.833	.543	.254			
3	90	DOS	.696	.769	.517		1	
4	90	FIP	.958	.581	.531	1	1	
5	90	BUB	.833	.637	.461	1		
6	90	HIC	.833	.817	.545		1	
1	146	SAX	.769	.774	.244			
2	146	JEN	.833	.543	.437	1		
3	146	TIB	.957	.487	.194			
4	146	NOV	.417	.610	.442		1	54% midvowel long
5	146	DAG	.875	.395	.513			
6	146	RAD	.875	.855	.455	1		

Legend:

CAT DISCR= discriminating power of item within category R101
TOT DISCR= discriminating power of item in total CVC test
CR TEST= item selected for criterion-referenced test (items chosen on basis of representing full range of item difficulties; CIT, TYG, COQ and NOV excluded because of error types)
NR TEST= item selected for norm-referenced test (items chosen on basis of a combination of medium difficulty and high discriminating power)

which best discriminate between student performances are not necessarily those which belong to a category (e.g. nov, coq in Table II). The criterion-referenced selection process has difficulties, too. Firstly, by removing items form the pool, the process of randomization is interfered with; although this may not be too harmful if items are selected on the basis of a range in item difficulty, rather than on the basis of medium difficulty. A larger threat is that by deciding to withdraw an item on the basis of its error type, one may be creating a new category. This encourages further and further subdivisions.

A combination of ordering theory and item analysis should result in being able to decide on relatively few items per category that (a) measure the trait and (b) represent the full range of difficulty. Ordering theory should also allow on a second try-out, for items to be ordered according to their hierarchies. If this is possible, then levels could be established on the basis of relatively few items, and a student could enter the test at a level appropriate to his ability.

It also suggests that there might be a case for the development of tests that use norm-referenced procedures for determining entry level, and criterion-referenced procedures for determining what a student knows or does not know at that level.

BIBLIOGRAPHY

Archer, E.J. A re-evaluation of the meaningfulness of all possible CVC trigrams. Psychological Monographs, 1960, 74 (10), Whole No. 497.

Bailey, M.H. The utility of phonic generalizations in grades one through six. The Reading Teacher, 1967, 20 (5), 413-418.

Bart, W.M. & Krus, D.J. An ordering-theoretic method to determine hierarchies among items. Educational and Psychological Measurement, 1973, 33, 291-300.

Bayles, E.E. Limitations of the Morrison unit. Science Education, 1934, 18, 203-207.

Bentler, P.M. Monotonicity analysis: an alternative to factor and test analysis. Paper presented at the Symposium on ordinal scales for development, Monterey, California, 1968.

Berdiansky, B., Cronnell, B. & Koehler, J. Spelling-sound relations and primary form-class descriptions for speech-comprehension vocabularies of 6-9 year olds. Technical Report #15, SWRL Educational Research and Development, California, June 1969.

Bissell, J.S. Planned variation in Head Start and Follow Through. In, J.S. Stanley (Ed). Pre-school programs for the disadvantaged, Volume II. Baltimore, Maryland: Johns Hopkins University Press, 1973.

Bloom, B.S., Hastings, T. & Madaus, G.F. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Bock, R.D. Multivariate statistical methods in behavioral research. Chicago: University of Chicago Press, 1974 (in press).

Bormuth, J.R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Bormuth, J.R. Development of standards of readability: toward a rational criterion of passage performance. Final Report, Grant #OEG-0-9-230237-4125 (010), University of Chicago, June 1971.

Brennan, R.L. Some statistical problems in the evaluation of self-instructional programs. Unpublished doctoral dissertation, Harvard University, Cambridge, Massachusetts, 1970.

Brown, R. A first language. Cambridge: Harvard University Press, 1973.

Bukovec, J.A. Usefulness of phonics generalizations: a new formula. The Reading Teacher, 1973, 27 (3), 270-274.

California Reading Test. J.D. Ayres, et. al. California: McGraw-Hill, 1936-1970.

Carroll, J.B., Davies, P. & Richman, B. (Eds). Word frequency book. Boston: Houghton Mifflin, 1971.

Chall, J.S. Learning to read: the great debate. New York: McGraw-Hill, 1967.

Chall, J.S. Stages of reading development: a preliminary statement of a study in progress. Paper presented at Harvard Summer Institute on Reading and Learning Disabilities, July 1973.

Chomsky, N. & Halle, M. The sound pattern of English. New York: Harper & Row, 1968.

Clymer, T. The utility of phonic generalizations in the primary grades. The Reading Teacher, 1963, 16 (4), 252-258.

Cooper-McGuire Diagnostic Word Analysis Test for Word Attack Skills. J.L. Cooper & M.L. McGuire. New London, Connecticut: Croft Educational Services, 1974.

Corder, R. The information base for reading: a critical review of the information base for current assumptions regarding the status of instruction and achievement in reading in the United States. Berkeley, California: Educational Testing Service, 1971.

Cox, R.C. Evaluative aspects of criterion-referenced measures. In, W.J. Popham (Ed). Criterion-referenced measurement: an introduction. Englewood Cliffs, N.J.: Educational Technology Publications, 1971.

Cox, R.C. & Sterrett, B.G. The application of a model for deriving more meaning from standardized test results. Working Paper #42, Learning R & D Center, University of Pittsburgh, 1968.

Cox, R.C. & Vargas, J.S. A comparison of item selection techniques for norm-referenced and criterion-referenced tests. Reprint #7, Learning Research and Development Center, University of Pittsburgh, 1966.

Cronbach, L.J. Test validation. In, R.L. Thorndike (Ed) Educational Measurement, 2nd Edition. Washington, D.C.: American Council on Education, 1971.

Cronbach, L.J., Rajaratnam, N. & Gleser, G.C. Theory of generalizability: a liberalization of reliability theory. British Journal of Statistical Psychology, 1963, 16, 137-163.

Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. The dependability of behavioral measurements: theory of generalizability for scores and profiles. New York: Wiley, 1972.

Dahl, T. Toward an evaluative methodology for criterion-referenced measures: objective-item congruence. Paper presented at California ERA, San Diego, 1971.

Davis, F.B. Fundamental factors of comprehension in reading. Psychometrica, 1944, 9, 185-197.

Davis, F.B. Criterion-referenced measurement. Paper presented at AERA Conference, New York, March 1972.

Davis, L.S. The applicability of phonics generalizations to selected spelling programs. Elementary English, 1972, 49, 706-713.

The Domain Phonics Test. J. McLeod & J. Atkinson. Edinburgh: Oliver & Boyd, 1972.

Durkin, D. Phonics, linguistics and reading. New York: Teacher's College Press, 1972.

Ebel, R.L. Essentials of educational measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Ebel, R.L. The future of the measurement of abilities--II. Educational Researcher, 1973, 2 (3), 5-12.

Emans, R. The usefulness of phonics generalizations above the primary grades. The Reading Teacher, 1967, 20 (5), 419-425.

Emans, R. & Harms, J.M. The usefulness of linguistically-based word generalizations. Reading World, 1973, 13 (1), 13-21.

Fries, C.C. Linguistics and reading. New York: Holt, Rinehart & Winston, 1963

Garvin, A.D. The applicability of criterion-referenced measurement by content area and level. Proceedings of the Symposium on Criterion-Referenced Measurement: Emerging Issues. Minneapolis, Minn., 1970.

Glaser, R. Instructional technology and the measurement of learning outcomes: some questions. American Psychologist, 1963, 18, 519-521.

Glaze, J.A. The association value of non-sense syllables. Journal of Genetic Psychology, 1928, 35, 255-267.

Goodman, K.S. (Ed). The psycholinguistic nature of the reading process. Detroit: Wayne State University Press, 1968.

Gray Oral Reading Test. W.S. Gray, et. al. New York: Bobbs-Merrill, 1963.

Gronland, N.E. Constructing achievement tests. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Hambleton, R.K. & Gorth, W.P. Criterion-referenced testing: issues and applications. Paper presented at the annual meeting of the NEERA, New York, 1970.

Hanna, P.R., Hanna, J.S., Hodges, R.E. & Rudorf, E.H. Phoneme-grapheme correspondences as cues to spelling improvement. Washington, D.C.: U.S. Government Printing Office, 1966.

Harris, C.W. An interpretation of Livingston's reliability coefficient for criterion-referenced tests. Journal of Educational Measurement, 1972, 9, 27-29.

Hawes, G.R. Criterion-referenced testing. Nation's Schools, 1973, 91, 35-41.

Henrysson, S. Gathering, analyzing, and using data on test items. In, R.L. Thorndike (Ed). Educational Measurement, 2nd Edition. Washington, D.C.: American Council on Education, 1971.

Hively, W., Patterson, H.L. & Page, S.H. A universe-defined system of arithmetic achievement tests. Journal of Educational Measurement, 1968, 5, 275-290.

Hopkins, K.D. & Wilkerson, C.J. Differential content validity: the California Spelling Test, an illustrative example. Educational and Psychological Measurement, 1965, 25, 413-419.

Hsu, Tse-Chi. Empirical data on criterion-referenced tests. Paper presented at the AERA meeting, New York, February 1971.

Inidividualized Criterion-referenced Testing--Reading. Tulsa, Oklahoma: Educational Progress, 1974.

Ivens, S.H. An investigation of item analysis, reliability and validity in relation to criterion-referenced tests. Unpublished doctoral dissertation, Florida State University, 1970.

Johnson, D.D. Suggested sequences for presenting four categories of letter-sound correspondences. Elementary English, 1973, 50 (6), 888-896.

Kerfoot, J.K. Comprehending comprehension. In, J.A. Figurel (Ed). Improvement of reading through classroom practice. International Reading Association Proceedings, Volume 9. Newark, Delaware, 1964.

Kirk, R.E. Experimental design: procedures for the behavioral sciences. Belmont, California: Wadsworth, 1968.

Kohlberg, L. Moral development. In, International Encyclopedia of Social Sciences. New York: Macmillan and Free Press, 1968.

Kohlberg, L. & Mayer, R. Development as the aim for education. Harvard Educational Review, 1972, 42 (4), 449-496.

Kreitler, S. & Kreitler, H. Dimensions of meaning and their measurement. Psychological Reports, 1968, 23, 1307-1329.

Kriewall, T.E. Applications of information theory and acceptance sampling principles to the management of mathematics instruction. Parts I and II. Technical Report #103. Wisconsin: Research & Development Center for Cognitive Learning, University of Wisconsin, 1969.

Krueger, W.C.F. The relative difficulty of nonsense syllables. Journal of Experimental Psychology, 1934, 17, 145-153.

Liberman, I.Y. Segmentation of the spoken word and reading acquisition. Bulletin of the Orton Society, 1973, 23, 65-77.

Livingston, S.A. The reliability of criterion-referenced measures. Report #73, Center for the Study of Social Organization of the Schools, Johns Hopkins University, July 1970.

Livingston, S.A. Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 1972, 9, 13-26.

Loevinger, J. Person and population as psychometric concepts. Psychological Review, 1965, 72 (2), 143-155.

Mann, L. & Phillips, W.A. Fractional practices in special education: a critique. Exceptional children, 1967, 33, 311-317.

McGraw-Hill Prescriptive Reading Inventory. Monterey, California: McGraw-Hill, 1972.

McGuire-Bumpus Diagnostic Test for Comprehension Skills. M.L. McGuire & M.J. Bumpus. New London, Connecticut: Croft Educational Services, 1974.

Metropolitan Achievement Test. B. Durrost, et. al. New York: Harcourt, Brace & World, 1931-1971.

Miller, G.A. Linguistic communication: perspectives for research. Newark, Delaware: International Reading Association, 1973.

Millman, J. Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 1973, 43 (2), 205-216.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Nunnally, J.C. Psychometric theory. New York: McGraw-Hill, 1967.

O'Reilly, R.B. & Hambleton, R.K. A CMI model for an individualized learning program in 9th grade science. Technical Report #14. Amherst: University of Massachusetts School of Education, 1971.

Osburn, H.G. Item sampling for achievement testing. Educational and Psychological Measurement, 1968, 28, 95-104.

Otis-Lennon Mental Ability Test. A.S. Otis & R.T. Lennon. New York: Harcourt, 1936-1970.

Ozenne, D.G. Toward an evaluative methodology for criterion-referenced measures: test sensitivity. Report #CSE-R-72, Center for the Study of Evaluation, California University, October 1971.

Pearson, K. Tables for statisticians and biometricians. Cambridge: Cambridge University Press, 1931.

Piaget, J. & Inhelder, B. The psychology of the child. New York: Basic Books, 1969.

Polanyi, M. Personal knowledge: towards a post-critical philosophy. Chicago: University of Chicago Press, 1958.

Popham, W.J. & Husek, T.R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Popp, H.M. The frequency of occurrence of graphemes in various positions (after Thorndike-Lorge, 1944). Harvard University, Cambridge, Massachusetts, 1963.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960.

Rasch, G. An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966, 19 (1), 49-57.

Shavelson, R., Block, J. & Ravitch, M. Criterion-referenced testing: comments on reliability. Journal of Educational Measurement, 1972, 9 (2), 133-137.

Sherwin, J.S. Four problems in teaching English: a critique of the research. Scranton, Penn.: International Textbook Company, 1969.

Shoemaker, D.M. Principles and procedures of multiple matrix sampling. Cambridge, Mass.: Ballinger, 1972.

Simon, D.P. & Simon, H.A. Alternative uses of phonemic information in spelling. Review of Educational Research, 1973, 43 (1), 115-137.

Simons, H.D. Reading comprehension: the need for a new perspective. Reading Research Quarterly, 1971, 6(3), 338-363.

Sipay Word Analysis Test. E. Sipay. Cambridge, Mass.: Educators Publishing, 1974.

Skinner, B.F. Why we need teaching machines. Harvard Educational Review, 1961, 31, 377-398.

Smith, F. Understanding reading. New York: Holt, Rinehart & Winston, 1971.

Spache, G.D. The teaching of reading. Bloomington, Indiana: Phi Delta Kappa, 1972.

Stanford Diagnostic Reading Test. B. Karlsen, et. al. New York: Harcourt, Brace & World, 1966-1968.

Stanley, J.C. Reliability. In, R.L. Thorndike (Ed). Educational Measurement, 2nd Edition. Washington, D.C.: American Council on Education, 1971.

Start, K. & Wells, B.K. The trend of reading standards. Slough, England: National Foundation for Educational Research in England & Wales, 1972.

Sticht, T.G. (Ed). Theory and method for the design of functional written materials (working draft). Monterey, California: Human Resources Reasearch Organization, 1974.

Thompson, R.A. & Dziubau, C.D. Criterion-referenced reading tests in perspective. The Reading Teacher, 1973, 27 (3), 292-294.

Thorndike, E.L. The measurement of ability in reading. Teacher's College Record, 1914, 15, 207-277.

Thorndike, E.L. & Lorge, I. The teacher's word book of 30,000 words. New York: Teachers College Press, 1944.

Thorndike, R.L. (Ed). Educational Measurement, 2nd Edition. Washington, D.C.: American Council on Education, 1971.

Thorndike, R.L. Reading as reasoning. Reading Research Quarterly, 1973, 2, 135-137.

Trismen, D. Adult readers: activities and goals. Paper presented at the meeting of the AERA, Chicago, 1972.

Tuinman, J.J. Determining the passage dependency on comprehension questions in five major tests. Reading Research Quarterly, 1973, 2, 206-223.

Underwood, B.J. & Shulz, R.W. Meaningfulness and verbal learning. Chicago: Lippincott, 1960.

Venezky, R.L. A computer program for deriving spelling to sound correlations. In, H. Levin (Ed). A basic research program on reading. Ithaca: Cornell University, 1963.

Venezky, R.L. The structure of English orthography. The Hague: Moulton, 1970.

Vernon, M.D. Reading and its difficulties. New York: Cambridge University Press, 1971.

Walmsley, S.A. Norm-referenced and criterion-referenced measurement and their relation to reading. Unpublished special qualifying paper, Harvard Graduate School of Education, Harvard University, Cambridge, Massachusetts, 1973.

Wardhaugh, R. Reading: a linguistic perspective. New York: Harcourt, Brace & World, 1969.

Weber, R. A linguistic analysis of first grade reading errors. Reading Research Quarterly, 1970, 5, 427-451.

Webster's International Dictionary. P.B. Grove (Ed). Springfield, Mass.: Merriam Company, 1972.

Wechsler Intelligence Scale for Children. D. Wechsler, et. al. New York: Psychological Corporation, 1949-1974.

Weir, R.H. & Venezky, R.L. A study of selected spelling to sound correspondence patterns. Stanford: Stanford University, 1966.

White, S. Some general outlines of the matrix of developmental changes between 5 and 7 years. Bulletin of the Orton Society, 1970, 20, 41-57.

Wright, B.D. Sample-free test calibration and person measurement. Chicago, Ill.: University of Chicago, 1967.

CVC UNIVERSE BY CATEGORIES--SEAN WALMSLEY, 1975

CVC	REAL WORD FREQ 00-6900	REAL WORD RARE, OBS SLANG	NONSENSE SYLLABLE	REAL WORD FREQ G1-4	INIT CONS FREQ	MED VOWEL FREQ	FINAL CONS FREQ	PRONUNCIATION PATTERN: PRON VAL	REG PRON	R MOD	W MOD	Y MOD	H MOD	ALT PRON	IRREG PRON	ASSOC VALUE	FREQ AS POLY SYLL	EXIST AS POLY SYLL	ALTERNATIVE PRONUNCIATION
MAB	X			0000	1	1	2	1	1							050	00	2	
MAC	X			0002	1	1	3	1	1					1		099	00	1	C11
MAD	X			0031	1	1	1	1	1							100	03	1	
MAF			X	-	1	1	3	1	1							026	00	1	
MAG	X			0000	1	1	3	1	1					1		095	04	1	G11
MAH			X	-	1	1	3	1					1			076	00	2	
MAJ			X	-	1	1	4	2	1							074	00	1	
MAK			X	-	1	1	2	1	1							087	00	2	
MAL		X		0000	1	1	2	1	1					1		094	07	1	A23
MAM		X		0000	1	1	2	1	1							-	04	1	
MAN	X			0978	1	1	1	1	1							099	29	1	
MAP	X			0193	1	1	2	1	1							100	00	1	
MAQ			X	-	1	1	3	2							1	038	-	-	
MAR	X			0000	1	1	1	1		1				1		099	13	1	A33
MAS		X		0000	1	1	3	1	1							091	03	1	
MAT	X			0021	1	1	1	1	1							100	03	1	
MAV			X	-	1	1	3	1	1							058	00	1	
MAW		X		0000	1	1	4	2			1					097	00	1	
MAX	X			0000	1	1	3	1	1							099	00	1	
MAY	X			1112	1	1	4	2				1				100	03	1	
MAZ			X	-	1	1	3	1	1							072	00	2	
MEB			X	-	1	3	2	1	1							033	-	-	
MEC			X	-	1	3	3	1	1							067	00	1	
MED			X	-	1	3	1	1	1							100	02	1	
MEF			X	-	1	3	3	1	1							022	-	-	
MEG	X			0000	1	3	3	1	1							095	02	1	
MEH			X	-	1	3	3	1					1		1	032	-	-	
MEJ			X	-	1	3	4	2	1							017	-	-	
MEK			X	-	1	3	2	1	1							053	-	-	
MEL	X			0000	1	3	2	1	1							097	03	1	
MEM			X	-	1	3	2	1	1							-	01	1	
MEN	X			0612	1	3	1	1	1							100	06	1	
MEP			X	-	1	3	2	1	1							036	-	-	
MEQ			X	-	1	3	4	2							1	027	-	-	
MER		X		0000	1	3	1	1		1				1		076	18	1	E33
MES			X	-	1	3	3	1	1					1		089	04	1	S20
MET	X			0081	1	3	1	1	1							100	04	1	
MEV			X	-	1	3	3	1	1							031	-	-	
MEW	X			0019	1	3	4	2			1					095	00	1	
MEX			X	-	1	3	3	1	1							067	00	1	
MEY			X	-	1	3	4	2				1				065	00	1	
MEZ			X	-	1	3	3	1	1							052	00	1	
MIB			X	-	1	1	2	1	1							027	-	-	
MIC			X	-	1	1	3	1	1					1		081	01	1	C11
MID	X			0001	1	1	1	1	1							099	06	1	
MIF			X	-	1	1	3	1	1							052	00	1	

APPENDIX (A), cont'd

CVC	REAL WORD FREQ 00-6900	REAL WORD RARE, OBS SLANG	NONSENSE SYLLABLE	REAL WORD FREQ G1-4	INIT CONS FREQ	MED VOWEL FREQ	FINAL CONS FREQ	PRONUNCIATION PATTERN: PRON VAL	REG PRON	R MOD	W MOD	Y MOD	H MOD	ALT PRON	IRREG PRON	ASSOC VALUE	FREQ AS POLY SYLL	EXIST AS POLY SYLL	ALTERNATIVE PRONUNCIATION
MAB	X			0000	1	1	2	1	1							050	00	2	
MAC	X			0002	1	1	3	1	1					1		099	00	1	C11
MAD	X			0031	1	1	1	1	1							100	03	1	
MAF			X	-	1	1	3	1	1							026	00	1	
MAG	X			0000	1	1	3	1	1					1		095	04	1	G11
MAH			X	-	1	1	3	1					1			076	00	2	
MAJ			X	-	1	1	4	2	1							074	00	1	
MAK			X	-	1	1	2	1	1							087	00	2	
MAL		X		0000	1	1	2	1	1					1		094	07	1	A23
MAM		X		0000	1	1	2	1	1							-	04	1	
MAN	X			0978	1	1	1	1	1							099	29	1	
MAP	X			0193	1	1	2	1	1							100	00	1	
MAQ			X	-	1	1	4	2							1	038	-	-	
MAR	X			0000	1	1	1	1		1				1		099	13	1	A33
MAS		X		0000	1	1	3	1	1							091	03	1	
MAT	X			0021	1	1	1	1	1							100	03	1	
MAV			X	-	1	1	3	1	1							058	00	1	
MAW		X		0000	1	1	4	2			1					097	00	1	
MAX	X			0000	1	1	3	1	1							099	00	1	
MAY	X			1112	1	1	4	2				1				100	03	1	
MAZ			X	-	1	1	3	1	1							072	00	2	
MEB			X	-	1	3	2	1	1							033	-	-	
MEC			X	-	1	3	3	1	1							067	00	1	
MED			X	-	1	3	1	1	1							100	02	1	
MEF			X	-	1	3	3	1	1							022	-	-	
MEG	X			0000	1	3	3	1	1							095	02	1	
MEH			X	-	1	3	3	1					1		1	032	-	-	
MEJ			X	-	1	3	4	2	1							017	-	-	
MEK			X	-	1	3	2	1	1							053	-	-	
MEL	X			0000	1	3	2	1	1							097	03	1	
MEM			X	-	1	3	2	1	1							-	01	1	
MEN	X			0612	1	3	1	1	1							100	06	1	
MEP			X	-	1	3	2	1	1							036	-	-	
MEQ			X	-	1	3	4	2							1	027	-	-	
MER		X		0000	1	3	1	1		1				1		076	18	1	E33
MES			X	-	1	3	3	1	1					1		089	04	1	S20
MET	X			0081	1	3	1	1	1							100	04	1	
MEV			X	-	1	3	3	1	1							031	-	-	
MEW	X			0019	1	3	4	2			1					095	00	1	
MEX			X	-	1	3	3	1	1							067	00	1	
MEY			X	-	1	3	4	2				1				065	00	1	
MEZ			X	-	1	3	3	1	1							052	00	1	
MIB			X	-	1	1	2	1	1							027	-	-	
MIC			X	-	1	1	3	1	1					1		081	01	1	C11
MID	X			0001	1	1	1	1	1							099	06	1	
MIF			X	-	1	1	3	1	1							052	00	1	